(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 875 570 A2

(12) # EUROPEAN PATENT APPLICATION

(43) Date of publication:
04.11.1998 Bulletin 1998/45

(51) Int Cl.6: **C12N 15/12**, C07K 14/47, C07K 16/18, G01N 33/68

(21) Application number: 98302912.5

(22) Date of filing: 15.04.1998

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: 01.05.1997 GB 9708936
18.12.1997 EP 97310289
19.02.1998 GB 9803566

(71) Applicant: **SMITHKLINE BEECHAM PLC**
Brentford, Middlesex TW8 9EP (GB)

(72) Inventors:
- **Harrison, David C.**
Harlow, Essex CM19 5AW (GB)
- **Davis, John**
Harlow, Essex CM19 5AW (GB)
- **Bingham, Sharon**
Harlow, Essex CM19 5AW (GB)
- **Doe, Trudy R.**
Harlow, Essex CM19 5AW (GB)
- **Topp, Simon**
Harlow, Essex CM19 5AW (GB)

(74) Representative: **Crump, Julian Richard John et al**
fJ Cleveland,
40-43 Chancery Lane
London WC2A 1JQ (GB)

(54) **Neurodegenerative polypeptides HHPDZ65**

(57) HHPDZ65 polypeptides and polynucleotides and methods for producing such polypeptides by recombinant techniques are disclosed. Also disclosed are methods for utilizing HHPDZ65 polypeptides and polynucleotides in therapy, and diagnostic assays for such.

EP 0 875 570 A2

Description

Field of the Invention

This invention relates to newly identified polypeptides and polynucleotides encoding such polypeptides, to their use in therapy and in identifying compounds which may be agonists, antagonists and /or inhibitors which are potentially useful in therapy, and to production of such polypeptides and polynucleotides.

Background of the Invention

The drug discovery process is currently undergoing a fundamental revolution as it embraces 'functional genomics', that is, high throughput genome- or gene-based biology. This approach as a means to identify genes and gene products as therapeutc targets is rapidly superceding earlier approaches based on 'positional cloning'. A phenotype, that is a biological function or genetic disease, would be identified and this would then be tracked back to the responsible gene, based on its genetic map position.

Functional genomics relies heavily on high-throughput DNA sequencing technologies and the various tools of bioinformatics to identify gene sequences of potential interest from the many molecular biology databases now available. There is a continuing need to identify and characterise further genes and their related polypeptides/proteins, as targets for drug discovery.

Summary of the Invention

The present invention relates to HHPDZ65, in particular HHPDZ65 polypeptides and HHPDZ65 polynucleotides, recombinant materials and methods for their production. In another aspect, the invention relates to methods for using such polypeptides and polynucleotides, including the treatment of stroke, pain, epilepsy, neurodegenerative diseases, hereinafter referred to as "the diseases", amongst others. In a further aspect, the invention relates to methods for identifying agonists and antagonists/inhibitors using the materials provided by the invention, and treating conditions associated with HHPDZ65 imbalance with the identified compounds. In a still further aspect, the invention relates to diagnostic assays for detecting diseases associated with inappropriate HHPDZ65 activity or levels.

Description of the Invention

In a first aspect, the present invention relates to HHPDZ65 polypeptides. Such peptides include isolated polypeptides comprising an amino acid sequence which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, most preferably at least 97-99% identity, to that of SEQ ID NO:2 or SEQ ID NO:8 over the entire length of SEQ ID NO:2 and SEQ ID NO:8. Such polypeptides include those comprising the amino acid sequences of SEQ ID NO:2 or SEQ ID NO:8. A comparison of the polypeptides of SEQ ID NO:2 (HHPDZ65) and SEQ ID NO:8 (HHPDZ65var) show that they share identical amino acids at 535 out of 539 positions. The 4 exceptions are:

Position 88 - proline in HHPDZ65 and leucine in HHPDZ65var
Position 506 - glutamine in HHPDZ65 and proline in HHPDZ65var
Position 508 - arginine in HHPDZ65 and leucine in HHPDZ65var
Position 511 - valine in HHPDZ65 and alanine in HHPDZ65var.

Further peptides of the present invention include isolated polypeptides in which the amino acid sequence has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, most preferably at least 97-99% identity, to the amino acid sequences of SEQ ID NO:2 or SEQ ID NO:8 over the entire length of SEQ ID NO:2 or SEQ ID NO:8. Such polypeptides include the polypeptides of SEQ ID NO:2 and SEQ ID NO:8.

Further peptides of the present invention include isolated polypeptides encoded by a polynucleotide comprising the sequence contained in SEQ ID NO: 1 or SEQ ID NO:7.

Polypeptides of the present invention are believed to be members of the degenerin family of polypeptides. They are therefore of interest because known members of this family cause neurodegeneration or are activated by ligands which are important in the above diseases. These properties are hereinafter referred to as HHPDZ65 activity" or HHPDZ65 polypeptide activity" or "biological activity of HHPDZ65". Also included amongst these activities are antigenic and immunogenic activities of said HHPDZ65 polypeptides, in particular the antigenic and immunogenic activities of the polypeptides of SEQ ID NO:2 or SEQ ID NO:8. Preferably, a polypeptide of the present invention exhibits at least

one biological activity of HHPDZ65.

The polypeptides of the present invention may be in the form of the "mature" protein or may be a part of a larger protein such as a precursor or a fusion protein. It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification such as multiple histidine residues, or an additional sequence for stability during recombinant production.

The present invention also includes variants of the aforementioned polypetides, that is polypeptides that vary from the referents by conservative amino acid substitutions, whereby a residue is substituted by another with like characteristics. Typical such substitutions are among Ala, Val, Leu and Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln; and among the basic residues Lys and Arg; or aromatic residues Phe and Tyr. Particularly preferred are variants in which several, 5-10, 1-5, 1-3, 1-2 or 1 amino acids are substituted, deleted, or added in any combination.

Polypeptides of the present invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

In a further aspect, the present invention relates to HHPDZ65 polynucleotides. Such polynucleotides include isolated polynucleotides comprising a nucleotide sequence encoding a polypeptide which has at least 75% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, to the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:8, over the entire length of SEQ ID NO:2 or SEQ ID NO:8. In this regard, polypeptides which have at least 97% identity are highly preferred, whilst those with at least 98-99% identity are more highly preferred, and those with at least 99% identity are most highly preferred. Such polynucleotides include a polynucleotide comprising the nucleotide sequence contained in SEQ ID NO:1 encoding the polypeptide of SEQ ID NO:2 and a polynucleotide comprising the nucleotide sequence contained in SEQ ID NO:7 encoding the polypeptide of SEQ ID NO:8.

Further polynucleotides of the present invention include isolated polynucleotides comprising a nucleotide sequence that has at least 75% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, to a nucleotide sequence encoding a polypeptide of SEQ ID NO:2 or SEQ ID NO:8, over the entire coding region. In this regard, polynucleotides which have at least 97% identity are highly preferred, whilst those with at least 98-99% identity are more highly preferred, and those with at least 99% identity are most highly preferred.

Further polynucleotides of the present invention include isolated polynucleotides comprising a nucleotide sequence which has at least 75% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, to SEQ ID NO: 1 or SEQ ID NO:7 over the entire length of SEQ ID NO: 1 or SEQ ID NO: 7. In this regard, polynucleotides which have at least 97% identity are highly preferred, whilst those with at least 98-99% identity are more highly preferred, and those with at least 99% identity are most highly preferred. Such polynucleotides include a polynucleotide comprising the polynucleotide of SEQ ID NO:1 or SEQ ID NO:7 as well as the polynucleotides of SEQ ID NO:1 and SEQ ID NO:7.

The invention also provides polynucleotides which are complementary to all the above described polynucleotides.

The nucleotide sequences of SEQ ID NO: 1 and SEQ ID NO:7 shows homology with human sodium channel 2, BNAC2 (Garcia -Anoveros, J et al, Proc Nat Acad Sci (1997) 94, 1459-1464). The nucleotide sequence of SEQ ID NO: 1 is a cDNA sequence and comprises a polypeptide encoding sequence (nucleotide 301 to 1917) encoding a polypeptide of 539 amino acids, the polypeptide of SEQ ID NO:2. The nucleotide sequence of SEQ ID NO:7 is a cDNA sequence and comprises a polypeptide encoding sequence (nucleotide 1 to 1617) encoding a polypeptide of 539 amino acids, the polypeptide of SEQ ID NO:8. The nucleotide sequence encoding the polypeptide of SEQ ID NO:2 may be identical to the polypeptide encoding sequence contained in SEQ ID NO: 1 or it may be a sequence other than the one contained in SEQ ID NO: 1, which, as a result of the redundancy (degeneracy) of the genetic code, also encodes the polypeptide of SEQ ID NO:2. Similarly the nucleotide sequence encoding the polypeptide of SEQ ID NO:8 may be identical to the polypeptide encoding sequence contained in SEQ ID NO:7 or it may be a sequence other than the one contained in SEQ ID NO:7, which, as a result of the redundancy (degeneracy) of the genetic code, also encodes the polypeptide of SEQ ID NO:8.

The polypeptides of SEQ ID NO:2 and SEQ ID NO:8 are structurally related to other proteins of the degenerin family, having homology and/or structural similarity with human sodium channel 2, BNAC2 (Garcia -Anoveros, J et al, Proc Nat Acad Sci (1997) 94, 1459-1464).

Preferred polypeptides and polynucleotides of the present invention are expected to have, *inter alia,* similar biological functions/properties to their homologous polypeptides and polynucleotides. Furthermore, preferred polypeptides and polynucleotides of the present invention have at least one HHPDZ65 activity.

The present invention also relates to partial or other polynucleotide and polypeptide sequences which were first identified prior to the determination of the corresponding full length sequences of SEQ ID NO:1 and SEQ ID NO:2.

Accordingly, in a further aspect, the present invention provides for an isolated polynucleotide comprising:

(a) a nucleotide sequence which has at least 75% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, even more preferably at least 97-99% identity to SEQ ID NO:3 over the entire length of SEQ ID NO:3;
(b) a nucleotide sequence which has at least 75% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, even more preferably at least 97-99% identity, to SEQ ID NO:1 over the entire length of SEQ ID NO:3;
(c) the polynucleotide of SEQ ID NO:3; or
(d) a nucleotide sequence encoding a polypeptide which has at least 75% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, even more preferably at least 97-99% identity, to the amino acid sequence of SEQ ID NO:4, over the entire length of SEQ ID NO:4;

as well as the polynucleotide of SEQ ID NO:3.

(e) a nucleotide sequence which has at least 80% identity, preferably at least 90% identity, more preferably at least 95% identity, even more preferably at least 97-99% identity to SEQ ID NO:5 over the entire length of SEQ ID NO:5;
(f) a nucleotide sequence which has at least 80% identity, preferably at least 90% identity, more preferably at least 95% identity, even more preferably at least 97-99% identity, to SEQ ID NO:1 over the entire length of SEQ ID NO:5;
(g) the polynucleotide of SEQ ID NO:5; or
(h) a nucleotide sequence encoding a polypeptide which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, even more preferably at least 97-99% identity, to the amino acid sequence of SEQ ID NO:6, over the entire length of SEQ ID NO:5;

as well as the polynucleotide of SEQ ID NO:5.

The present invention further provides for a polypeptide which:

(a) comprises an amino acid sequence which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, most preferably at least 97-99% identity, to that of SEQ ID NO:4 over the entire length of SEQ ID NO:4;
(b) has an amino acid sequence which is at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, most preferably at least 97-99% identity, to the amino acid sequence of SEQ ID NO:4 over the entire length of SEQ ID NO:4;
(c) comprises the amino acid of SEQ ID NO:4; and
(d) is the polypeptide of SEQ ID NO:4;

as well as polypeptides encoded by a polynucleotide comprising the sequence contained in SEQ ID NO:3.

(e) comprises an amino acid sequence which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, most preferably at least 97-99% identity, to that of SEQ ID NO:6 over the entire length of SEQ ID NO:6;
(f) has an amino acid sequence which is at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, most preferably at least 97-99% identity, to the amino acid sequence of SEQ ID NO:6 over the entire length of SEQ ID NO:6;
(g) comprises the amino acid of SEQ ID NO:6; and
(h) is the polypeptide of SEQ ID NO:6;

as well as polypeptides encoded by a polynucleotide comprising the sequence contained in SEQ ID NO:5.

The nucleotide sequences of SEQ ID NO:3 and SEQ ID NO: 5 and the peptide sequences encoded thereby are derived from EST (Expressed Sequence Tag) sequences. It is recognised by those skilled in the art that there will inevitably be some nucleotide sequence reading errors in EST sequences (see Adams, M.D. *et al*, Nature 377 (supp) 3, 1995). Accordingly, the nucleotide sequences of SEQ ID NO:3 and SEQ ID NO:5 and the peptide sequences encoded therefrom are therefore subject to the same inherent limitations in sequence accuracy. Furthermore, the peptide sequences encoded by SEQ ID NO:3 and SEQ ID NO:5 comprise a region of identity or close homology and/or close structural similarity (for example a conservative amino acid difference) with the closest homologous or structurally similar protein.

Polynucleotides of the present invention may be obtained, using standard cloning and screening techniques, from a cDNA library derived from mRNA in cells of human hippocampus, cerebellum, foetal brain, using the expressed sequence tag (EST) analysis (Adams, M.D., *et al.* Science (1991) 252:1651-1656; Adams, M.D. *et al.*, Nature, (1992) *355*:632-634; Adams, M.D., *et al.*, Nature (1995) 377 Supp:3-174). Polynucleotides of the invention can also be ob-

tained from natural sources such as genomic DNA libraries or can be synthesized using well known and commercially available techniques.

When polynucleotides of the present invention are used for the recombinant production of polypeptides of the present invention, the polynucleotide may include the coding sequence for the mature polypeptide, by itself; or the coding sequence for the mature polypeptide in reading frame with other coding sequences, such as those encoding a leader or secretory sequence, a pre-, or pro- or prepro- protein sequence, or other fusion peptide portions. For example, a marker sequence which facilitates purification of the fused polypeptide can be encoded. In certain preferred embodiments of this aspect of the invention, the marker sequence is a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz *et al.,* Proc Natl Acad Sci USA (1989) 86:821-824, or is an HA tag. The polynucleotide may also contain non-coding 5' and 3' sequences, such as transcribed, non-translated sequences, splicing and polyadenylation signals, ribosome binding sites and sequences that stabilize mRNA.

Further embodiments of the present invention include polynucleotides encoding polypeptide variants which comprise the amino acid sequences of SEQ ID NO:2 and SEQ ID NO:8 and in which several, for instance from 5 to 10, 1 to 5, 1 to 3, 1 to 2 or 1, amino acid residues are substituted, deleted or added, in any combination.

Polynucleotides which are identical or sufficiently identical to a nucleotide sequence contained in SEQ ID NO:1 or SEQ ID NO:7, may be used as hybridization probes for cDNA and genomic DNA or as primers for a nucleic acid amplification (eg. PCR) reaction, to isolate full-length cDNAs and genomic clones encoding polypeptides of the present invention and to isolate cDNA and genomic clones of other genes (including genes encoding paralogs from human sources and orthologs and paralogs from species other than human) that have a high sequence similarity to SEQ ID NO: 1 or SEQ ID NO:7. Typically these nucleotide sequences are 70% identical, preferably 80% identical, more preferably 90% identical, most preferably 95% identical to that of the referent. The probes or primers will generally comprise at least 15 nucleotides, preferably, at least 30 nucleotides and may have at least 50 nucleotides. Particularly preferred probes will have between 30 and 50 nucleotides. Particularly preferred primers will have between 20 and 25 nucleotides.

A polynucleotide encoding a polypeptide of the present invention, including homologs from species other than human, may be obtained by a process which comprises the steps of screening an appropriate library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO:1, SEQ ID NO:7 or fragments thereof; and isolating full-length cDNA and genomic clones containing said polynucleotide sequence. Such hybridization techniques are well known to the skilled artisan. Preferred stringent hybridization conditions include overnight incubation at 42°C in a solution comprising: 50% formamide, 5xSSC (150mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10 % dextran sulfate, and 20 microgram/ml denatured, sheared salmon sperm DNA; followed by washing the filters in 0.1x SSC at about 65°C. Thus the present invention also includes polynucleotides obtainable by screening an appropriate library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO:1, SEQ ID NO:7 or fragments thereof.

The skilled artisan will appreciate that, in many cases, an isolated cDNA sequence will be incomplete, in that the region coding for the polypeptide is short at the 5' end of the cDNA. This is a consequence of reverse transcriptase, an enzyme with inherently low 'processivity' (a measure of the ability of the enzyme to remain attached to the template during the polymerisation reaction), failing to complete a DNA copy of the mRNA template during 1st strand cDNA synthesis.

There are several methods available and well known to those skilled in the art to obtain full-length cDNAs, or extend short cDNAs, for example those based on the method of Rapid Amplification of cDNA ends (RACE) (see, for example, Frohman et al., PNAS USA 85, 8998-9002, 1988). Recent modifications of the technique, exemplified by the Marathon™' technology (Clontech Laboratories Inc.) for example, have significantly simplified the search for longer cDNAs. In the Marathon™ technology, cDNAs have been prepared from mRNA extracted from a chosen tissue and an 'adaptor' sequence ligated onto each end. Nucleic acid amplification (PCR) is then carried out to amplify the 'missing' 5' end of the cDNA using a combination of gene specific and adaptor specific oligonucleotide primers. The PCR reaction is then repeated using 'nested' primers, that is, primers designed to anneal within the amplified product (typically an adaptor specific primer that anneals further 3' in the adaptor sequence and a gene specific primer that anneals further 5' in the known gene sequence). The products of this reaction can then be analysed by DNA sequencing and a full-length cDNA constructed either by joining the product directly to the existing cDNA to give a complete sequence, or carrying out a separate full-length PCR using the new sequence information for the design of the 5' primer.

Recombinant polypeptides of the present invention may be prepared by processes well known in the art from genetically engineered host cells comprising expression systems. Accordingly, in a further aspect, the present invention relates to expression systems which comprise a polynucleotide or polynucleotides of the present invention, to host cells which are genetically engineered with such expression sytems and to the production of polypeptides of the invention by recombinant techniques. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention.

For recombinant production, host cells can be genetically engineered to incorporate expression systems or portions thereof for polynucleotides of the present invention. Introduction of polynucleotides into host cells can be effected by

methods described in many standard laboratory manuals, such as Davis et al., Basic Methods in Molecular Biology (1986) and Sambrook *et al.,* Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989). Preferred such methods include, for instance, calcium phosphate transfection, DEAE-dextran mediated transfections, transfection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction or infection.

Representative examples of appropriate hosts include bacterial cells, such as *Streptococci, Staphylococci, E. coli, Streptomyces* and *Bacillus subtilis* cells: fungal cells, such as yeast cells and *Aspergillus* cells: insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells: animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, HEK 293 and Bowes melanoma cells: and plant cells.

A great variety of expression systems can be used, for instance, chromosomal, episomal and virus-derived systems, e.g., vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression systems may contain control regions that regulate as well as engender expression. Generally, any system or vector which is able to maintain, propagate or express a polynucleotide to produce a polypeptide in a host may be used. The appropriate nucleotide sequence may be inserted into an expression system by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook *et al.,* Molecular Cloning, A Laboratory Manual (supra). Appropriate secretion signals may be incorporated into the desired polypeptide to allow secretion of the translated protein into the lumen of the endoplasmic reticulum, the periplasmic space or the extracellular environment. These signals may be endogenous to the polypeptide or they may be heterologous signals.

If polypeptide of the present invention is to be expressed for use in screening assays, it is generally preferred that the polypeptide be produced at the surface of the cell. In this event, the cells may be harvested prior to use in the screening assay. If the polypeptide is secreted into the medium, the medium can be recovered in order to recover and purify the polypeptide. If produced intracellularly, the cells must first be lysed before the polypeptide is recovered.

Polypeptides of the present invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography is employed for purification. Well known techniques for refolding proteins may be employed to regenerate active conformation when the polypeptide is denatured during intracellular synthesis, isolation and or purification.

This invention also relates to the use of polynucleotides of the present invention as diagnostic reagents. Detection of a mutated form of the gene characterised by the polynucleotides of SEQ ID NO: 1 or SEQ ID NO:7 which is associatcd with a dysfunction will provide a diagnostic tool that can add to, or define, a diagnosis of a disease, or susceptibility to a disease, which results from under-expression, over-expression or altered spatial or temporal expression of the gene. Individuals carrying mutations in the gene may be detected at the DNA level by a variety of techniques.

Nucleic acids for diagnosis may be obtained from a subject's cells, such as from blood, urine, saliva, tissue biopsy or autopsy material. The genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR or other amplification techniques prior to analysis. RNA or cDNA may also be used in similar fashion. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to labeled HHPDZ65 nucleotide sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase digestion or by differences in melting temperatures. DNA sequence differences may also be detected by alterations in electrophoretic mobility of DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing (ee, e.g., Myers *et al.,* Science (1985) 230:1242). Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method (see Cotton *et al.,* Proc Natl Acad Sci USA (1985) 85: 43974401). In another embodiment, an array of oligonucleotides probes comprising HHPDZ65 nucleotide sequence or fragments thereof can be constructed to conduct efficient screening of e.g., genetic mutations. Array technology methods are well known and have general applicability and can be used to address a variety of questions in molecular genetics including gene expression, genetic linkage, and genetic variability (see for example: M.Chee et al., Science, Vol 274, pp 610-613 (1996)).

The diagnostic assays offer a process for diagnosing or determining a susceptibility to the diseases through detection of mutation in the HHPDZ65 gene by the methods described. In addition, such diseases may be diagnosed by methods comprising determining from a sample derived from a subject an abnormally decreased or increased level of polypeptide or mRNA. Decreased or increased expression can be measured at the RNA level using any of the methods well known in the art for the quantitation of polynucleotides, such as, for example, nucleic acid amplification, for instance PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods. Assay techniques that can be used to determine levels of a protein, such as a polypeptide of the present invention, in a sample derived from a host

are well-known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays.

Thus in another aspect, the prcsent invention relates to a diagonostic kit which comprises:

(a) a polynucleotide of the present invention, preferably the nucleotide sequence of SEQ ID NO: 1, SEQ ID NO:7 or a fragment thereof:

(b) a nucleotide sequence complementary to that of (a);

(c) a polypeptide of the present invention, preferably the polypeptide of SEQ ID NO:2, SEQ ID NO:8 or a fragment thereof; or

(d) an antibody to a polypeptide of the present invention, preferably to the polypeptide of SEQ ID NO:2 or SEQ ID NO:8.

It will be appreciated that in any such kit, (a), (b), (c) or (d) may comprise a substantial component. Such a kit will be of use in diagnosing a disease or suspectability to a disease, particularly stroke, pain, epilepsy, neurodegenerative diseases, amongst others.

The nucleotide sequences of the present invention are also valuable for chromosome localisation. The sequence is specifically targeted to, and can hybridize with, a particular location on an individual human chromosome. The mapping of relevant sequences to chromosomes according to the present invention is an important first step in correlating those sequences with gene associated disease. Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found in, for example, V. McKusick, Mendelian Inheritance in Man (available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes).

The differences in the cDNA or genomic sequence between affected and unaffected individuals can also be determined. If a mutation is observed in some or all of the affected individuals but not in any normal individuals, then the mutation is likely to be the causative agent of the disease.

The nucleotide sequences of the present invention are also valuable for tissue localisation. Such techniques allow the determination of expression patterns of the HHPDZ65 polypeptides in tissues by detection of the mRNAs that encode them. These techniques include in situ hybridziation techniques and nucleotide amplification techniques, for example PCR. Such techniques are well known in the art. Results from these studies provide an indication of the normal functions of the polypeptides in the organism. In addition, comparative studies of the normal expression pattern of HHPDZ65 mRNAs with that of mRNAs encoded by a mutant HHPDZ65 gene provide valuable insights into the role of mutant HHPDZ65 polypeptides, or that of inappropriate expression of normal HHPDZ65 polypeptides, in disease. Such inappropriate expression may be of a temporal, spatial or simply quantitative nature.

The HHPDZ65 polypeptides of the invention are expressed in a wide range of tissues of the central nervous system, for example expression was detected in superior frontal gyrus, cerebellum, parahippocampal gyrus, caudate nucleus, amygdala, inferior temporal gyrus, hippocampus, medulla oblongata, nucleus acumbens, substantia nigra, striatum, superior occipital gyrus, hypothalamus, thalamus, putamen, spinal cord. The expression detected was highest in cerebellum, parahippocampal gyrus, caudate nucleus, amygdala, superior occipital gyrus and thalamus.

The polypeptides of the invention or their fragments or analogs thereof, or cells expressing them, can also be used as immunogens to produce antibodies immunospecific for polypeptides of the present invention. The term "immunospecific" means that the antibodies have substantially greater affinity for the polypeptides of the invention than their affinity for other related polypeptides in the prior art.

Antibodies generated against polypeptides of the present invention may be obtained by administering the polypeptides or epitope-bearing fragments, analogs or cells to an animal, preferably a non-human animal, using routine protocols. For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler, G. and Milstein, C., Nature (1975) 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor *et al.*, Immunology Today (1983) 4:72) and the EBV-hybridoma technique (Cole *et al.*, Monoclonal Antibodies and Cancer Therapy, 77-96, Alan R. Liss, Inc., 1985).

Techniques for the production of single chain antibodies, such as those described in U.S. Patent No. 4,946,778, can also be adapted to produce single chain antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms, including other mammals, may be used to express humanized antibodies.

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptide or to purify the polypeptides by affinity chromatography.

Antibodies against polypeptides of the present invention may also be employed to treat the diseases, amongst others.

In a further aspect, the present invention relates to genetically engineered soluble fusion proteins comprising a

polypeptide of the present invention, or a fragment thereof, and various portions of the constant regions of heavy or light chains of immunoglobulins of various subclasses (IgG, IgM, IgA, IgE). Preferred as an immunoglobulin is the constant part of the heavy chain of human IgG, particularly IgG1, where fusion takes place at the hinge region. In a particular embodiment, the Fc part can be removed simply by incorporation of a cleavage sequence which can be cleaved with blood clotting factor Xa. Furthermore, this invention relates to processes for the preparation of these fusion proteins by genetic engineering, and to the use thereof for drug screening, diagnosis and therapy. A further aspect of the invention also relates to polynucleotides encoding such fusion proteins. Examples of fusion protein technology can be found in International Patent Application Nos. WO94/29458 and WO94/22914.

Another aspect of the invention relates to a method for inducing an immunological response in a mammal which comprises inoculating the mammal with a polypeptide of the present invention, adequate to produce antibody and/or T cell immune response to protect said animal from the diseases hereinbefore mentioned, amongst others. Yet another aspect of the invention relates to a method of inducing immunological response in a mammal which comprises, delivering a polypeptide of the present invention *via* a vector directing expression of the polynucleotide and coding for the polypeptide *in vivo* in order to induce such an immunological response to produce antibody to protect said animal from diseases.

A further aspect of the invention relates to an immunological/vaccine formulation (composition) which, when introduced into a mammalian host, induces an immunological response in that mammal to a polypeptide of the present invention wherein the composition comprises a polypeptide or polynucleotide of the present invention. The vaccine formulation may further comprise a suitable carrier. Since a polypeptide may be broken down in the stomach, it is preferably administered parenterally (for instance, subcutaneous, intramuscular, intravenous, or intradermal injection). Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation instonic with the blood of the recipient; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampoules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in water systems and other systems known in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

Polypeptides of the present invention are responsible for one or more biological functions, including one or more disease states, in particular the diseases hereinbefore mentioned. It is therefore desirous to devise screening methods to identify compounds which stimulate or which inhibit the function of the polypeptide. Accordingly, in a further aspect, the present invention provides for a method of screening compounds to identify those which stimulate or which inhibit the function of the polypeptide. In general, agonists or antagonists may be employed for therapeutic and prophylactic purposes for such diseases as hereinbefore mentioned. Compounds may be identified from a variety of sources, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures. Such agonists, antagonists or inhibitors so-identified may be natural or modified substrates, ligands, receptors, enzymes, etc., as the case may be, of the polypeptide: or may be structural or functional mimetics thereof (see Coligan *et al.*, Current Protocols in Immunology 1(2):Chapter 5 (1991)).

The screening method may simply measure the binding of a candidate compound to the polypeptide, or to cells or membranes bearing the polypeptide, or a fusion protein thereof by means of a label directly or indirectly associated with the candidate compound. Alternatively, the screening method may involve competition with a labeled competitor. Further, these screening methods may test whether the candidate compound results in a signal generated by activation or inhibition of the polypeptide, using detection systems appropriate to the cells bearing the polypeptide. Inhibitors of activation are generally assayed in the presence of a known agonist and the effect on activation by the agonist by the presence of the candidate compound is observed. Constitutively active polypeptides may be employed in screening methods for inverse agonists or inhibitors, in the absence of an agonist or inhibitor, by testing whether the candidate compound results in inhibition of activation of the polypeptide. Further, the screening methods may simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide of the present invention, to form a mixture, measuring HHPDZ65 activity in the mixture, and comparing the HHPDZ65 activity of the mixture to a standard. Fusion proteins, such as those made from Fc portion and HHPDZ65 polypeptide, as hereinbefore described, can also be used for high-throughput screening assays to identify antagonists for the polypeptide of the present invention (see D. Bennett *et al.*, J Mol Recognition, 8:52-58 (1995); and K. Johanson *et al.*, J Biol Chem, 270(16):9459-9471 (1995)).

The polynucleotides, polypeptides and antibodies to the polypeptide of the present invention may also be used to configure screening methods for detecting the effect of added compounds on the production of mRNA and polypeptide in cells. For example, an ELISA assay may be constructed for measuring secreted or cell associated levels of polypeptide using monoclonal and polyclonal antibodies by standard methods known in the art. This can be used to discover agents which may inhibit or enhance the production of polypeptide (also called antagonist or agonist, respectively) from suitably manipulated cells or tissues.

The polypeptide may be used to identify membrane bound or soluble receptors, if any, through standard receptor binding techniques known in the art. These include, but are not limited to, ligand binding and crosslinking assays in which the polypeptide is labeled with a radioactive isotope (for instance, ^{125}I), chemically modified (for instance, biotinylated), or fused to a peptide sequence suitable for detection or purification, and incubated with a source of the putative receptor (cells, cell membranes, cell supernatants, tissue extracts, bodily fluids). Other methods include biophysical techniques such as surface plasmon resonance and spectroscopy. These screening methods may also be used to identify agonists and antagonists of the polypeptide which compete with the binding of the polypeptide to its receptors, if any. Standard methods for conducting such assays are well understood in the art.

Examples of potential polypeptide antagonists include antibodies or, in some cases, oligonucleotides or proteins which are closely related to the ligands, substrates, receptors, enzymes, etc., as the case may be, of the polypeptide, e.g., a fragment of the ligands, substrates, receptors, enzymes, etc.; or small molecules which bind to the polypeptide of the present invention but do not elicit a response, so that the activity of the polypeptide is prevented.

Thus, in another aspect, the present invention relates to a screening kit for identifying agonists, antagonists, ligands, receptors, substrates, enzymes, etc. for polypeptides of the present invention; or compounds which decrease or enhance the production of such polypeptides, which comprises:

(a) a polypeptide of the present invention;
(b) a recombinant cell expressing a polypeptide of the present invention;
(c) a cell membrane expressing a polypeptide of the present invention; or
(d) antibody to a polypeptide of the present invention;

which polypeptide is preferably that of SEQ ID NO:2 or SEQ ID NO:8.

It will be appreciated that in any such kit, (a), (b), (c) or (d) may comprise a substantial component.

It will be readily appreciated by the skilled artisan that a polypeptide of the present invention may also be used in a method for the structure-based design of an agonist, antagonist or inhibitor of the polypeptide, by:

(a) determining in the first instance the three-dimensional structure of the polypeptide;
(b) deducing the three-dimensional structure for the likely reactive or binding site(s) of an agonist, antagonist or inhibitor;
(c) synthesing candidate compounds that are predicted to bind to or react with the deduced binding or reactive site; and
(d) testing whether the candidate compounds are indeed agonists, antagonists or inhibitors.

It will be further appreciated that this will normally be an iterative process.

In a further aspect, the present invention provides methods of treating abnormal conditions such as, for instance, stroke, pain, epilepsy, neurodegenerative diseases, related to either an excess of, or an under-expression of, HHPDZ65 polypeptide activity.

If the activity of the polypeptide is in excess, several approaches are available. One approach comprises administering to a subject in need thereof an inhibitor compound (antagonist) as hereinabove described, optionally in combination with a pharmaceutically acceptable carrier, in an amount effective to inhibit the function of the polypeptide, such as, for example, by blocking the binding of ligands, substrates, receptors, enzymes, etc., or by inhibiting a second signal, and thereby alleviating the abnormal condition. In another approach, soluble forms of the polypeptides still capable of binding the ligand, substrate, enzymes, receptors, etc. in competition with endogenous polypeptide may be administered. Typical examples of such competitors include fragments of the HHPDZ65 polypeptide.

In still another approach, expression of the gene encoding endogenous HHPDZ65 polypeptide can be inhibited using expression blocking techniques. Known such techniques involve the use of antisense sequences, either internally generated or externally administered (see, for example, O'Connor, J Neurochem (1991) 56:560 in Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Alternatively, oligonucleotides which form triple helices ("triplexes") with the gene can be supplied (see, for example, Lee *et al.*, Nucleic Acids Res (1979) 6:3073; Cooney *et al.*, Science (1988) 241:456; Dervan *et al.*, Science (1991) 251:1360). These oligomers can be administered *per se* or the relevant oligomers can be expressed *in vivo*. Synthetic antisense or triplex oligonucleotides may comprise modified bases or modified backbones. Examples of the latter include methylphosphonate, phosphorothioate or peptide nucleic acid backbones. Such backbones are incorporated in the antisense or triplex oligonucleotide in order to provide protection from degradation by nucleases and are well known in the art. Antisense and triplex molecules synthesised with these or other modified backbones also form part of the present invention.

In addition, expression of the HHPDZ65 polypeptide may be prevented by using ribozymes specific to the HHPDZ65 mRNA sequence. Ribozymes are catalytically active RNAs that can be natural or synthetic (see for example Usman, N, et al., Curr. Opin. Struct. Biol (1996) 6(4), 527-33.) Synthetic ribozymes can be designed to specifically

cleave HHPDZ65 mRNAs at selected positions thereby preventing translation of the HHPDZ65 mRNAs into functional polypeptide. Ribozymes may be synthesised with a natural ribose phosphate backbone and natural bases, as normally found in RNA molecules. Alternatively the ribosymes may be synthesised with non-natural backbones to provide protection from ribonuclease degradation, for example, 2'-O-methyl RNA, and may contain modified bases.

For treating abnormal conditions related to an under-expression of HHPDZ65 and its activity, several approaches are also available. One approach comprises administering to a subject a therapeutically effective amount of a compound which activates a polypeptide of the present invention, i.e., an agonist as described above, in combination with a pharmaceutically acceptable carrier, to thereby alleviate the abnormal condition. Alternatively, gene therapy may be employed to effect the endogenous production of HHPDZ65 by the relevant cells in the subject. For example, a polynucleotide of the invention may be engineered for expression in a replication defective retroviral vector, as discussed above. The retroviral expression construct may then be isolated and introduced into a packaging cell transduced with a retroviral plasmid vector containing RNA encoding a polypeptide of the present invention such that the packaging cell now produces infectious viral particles containing the gene of interest. These producer cells may be administered to a subject for engineering cells *in vivo* and expression of the polypeptide *in vivo*. For an overview of gene therapy, see Chapter 20, Gene Therapy and other Molecular Genetic-based Therapeutic Approaches, (and references cited therein) in Human Molecular Genetics, T Strachan and A P Read, BIOS Scientific Publishers Ltd (1996). Another approach is to administer a therapeutic amount of a polypeptide of the present invention in combination with a suitable pharmaceutical carrier.

In a further aspect, the present invention provides for pharmaceutical compositions comprising a therapeutically effective amount of a polypeptide, such as the soluble form of a polypeptide of the present invention, agonist/antagonist peptide or small molecule compound, in combination with a pharmaceutically acceptable carrier or excipient. Such carriers include, but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol, and combinations thereof. The invention further relates to pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention. Polypeptides and other compounds of the present invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds.

The composition will be adapted to the route of administration, for instance by a systemic or an oral route. Preferred forms of systemic administration include injection, typically by intravenous injection. Other injection routes, such as subcutaneous, intramuscular, or intraperitoneal, can be used. Alternative means for systemic administration include transmucosal and transdermal administration using penetrants such as bile salts or fusidic acids or other detergents. In addition, if a polypeptide or other compounds of the present invention can be formulated in an enteric or an encapsulated formulation, oral administration may also be possible. Administration of these compounds may also be topical and/or localized, in the form of salves, pastes, gels, and the like.

The dosage range required depends on the choice of peptide or other compounds of the present invention, the route of administration, the nature of the formulation, the nature of the subject's condition, and the judgment of the attending practitioner. Suitable dosages, however, are in the range of 0.1-100 μg/kg of subject. Wide variations in the needed dosage, however, are to be expected in view of the variety of compounds available and the differing efficiencies of various routes of administration. For example, oral administration would be expected to require higher dosages than administration by intravenous injection. Variations in these dosage levels can be adjusted using standard empirical routines for optimization, as is well understood in the art.

Polypeptides used in treatment can also be generated endogenously in the subject, in treatment modalities often referred to as "gene therapy" as described above. Thus, for example, cells from a subject may be engineered with a polynucleotide, such as a DNA or RNA, to encode a polypeptide *ex vivo*, and for example, by the use of a retroviral plasmid vector. The cells are then introduced into the subject

Polynucleotide and polypeptide sequences form a valuable information resource with which to identify further sequences of similar homology. This is most easily facilitated by storing the sequence in a computer readable medium and then using the stored data to search a sequence database using well known searching tools, such as those in the GCG and Lasergene software packages. Accordingly, in a further aspect, the present invention provides for a computer readable medium having stored thereon a polynucleotide comprising the sequences of SEQ ID NO: 1 or SEQ ID NO: 7 and/or a polypeptide sequence encoded thereby.

The following definitions are provided to facilitate understanding of certain terms used frequently hereinbefore.

"Antibodies" as used herein includes polyclonal and monoclonal antibodies, chimeric, single chain, and humanized antibodies, as well as Fab fragments, including the products of an Fab or other immunoglobulin expression library.

"Isolated" means altered "by the hand of man" from the natural state. If an "isolated" composition or substance occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living animal is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein.

"Polynucleotide" generally refers to any polyribonucleotide or polydeoxribonucleotide, which may be unmodified

RNA or DNA or modified RNA or DNA. "Polynucleotides" include, without limitation, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, "polynucleotide" refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The term "polynucleotide" also includes DNAs or RNAs containing one or more modified bases and DNAs or RNAs with backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications may be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically or metabolically modified forms of polynucleotides as typically found in nature, as well as the chemical forms of DNA and RNA characteristic of viruses and cells. "Polynucleotide" also embraces relatively short polynucleotides, often referred to as oligonucleotides.

"Polypeptide" refers to any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres. "Polypeptide" refers to both short chains, commonly referred to as peptides, oligopeptides or oligomers, and to longer chains, generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene-encoded amino acids. "Polypeptides" include amino acid sequences modified either by natural processes, such as post-translational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications may occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present to the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched and branched cyclic polypeptides may result from post-translation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, biotinylation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination (see, for instance, Proteins - Structure and Molecular Properties, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York, 1993; Wold, F., Post-translational Protein Modifications: Perspectives and Prospects, pgs. 1-12 in Post-translational Covalent Modification of Proteins, B. C. Johnson, Ed., Academic Press, New York, 1983; Seifter *et al.*, "Analysis for protein modifications and nonprotein cofactors", Meth Enzymol (1990) 182: 626-646 and Rattan *et al.*, "Protein Synthesis: Post-translational Modifications and Aging", Ann NY Acad Sci (1992) 663:48-62).

"Variant" refers to a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide, but retains essential properties. A typical variant of a polynucleotide differs in nucleotide sequence from another, reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. A variant of a polynucleotide or polypeptide may be a naturally occurring such as an allelic variant, or it may be a variant that is not known to occur naturally. Non-naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques or by direct synthesis.

"Identity," as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. "Identity" and "similarity" can be readily calculated by known methods, including but not limited to those described in (Computational Molecular Biology, Lesk, A.M., ed., Oxford University Press, New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D.W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., SIAM *J. Applied Math., 48:* 1073 (1988). Preferred methods to determine identity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Preferred

computer program methods to determine identity and similarity between two sequences include, but are not limited to, the GCG program package (Devereux, J., et al., Nucleic Acids Research *12(1):* 387 (1984)), BLASTP, BLASTN, and FASTA (Atschul, S.F. et al., J. Molec. Biol. *215:* 403-410 (1990). The BLAST X program is publicly available from NCBI and other sources (BLAST Manual, Altschul, S., *et al.,* NCBI NLM NIH Bethesda, MD 20894: Altschul, S., *et al.,* J. Mol. Biol. *215:* 403-410 (1990). The well known Smith Waterman algorithm may also be used to determine identity.

Preferred parameters for polypeptide sequence comparison include the following:

1) Algorithm: Needleman and Wunsch, J. Mol Biol. 48: 443-453 (1970)

Comparison matrix: BLOSSUM62 from Hentikoff and Hentikoff, Proc. Natl. Acad. Sci. USA. 89:10915-10919 (1992)
Gap Penalty: 12
Gap Length Penalty: 4

A program useful with these parameters is publicly available as the "gap" program from Genetics Computer Group, Madison WI. The aforementioned parameters are the default parameters for polypeptide comparisons (along with no penalty for end gaps).

Preferred parameters for polynucleotide comparison include the following:

1) Algorithm: Needleman and Wunsch, J. Mol Biol. 48: 443-453 (1970)

Comparison matrix: matches = +10, mismatch = 0
Gap Penalty: 50
Gap Length Penalty: 3

A program useful with these parameters is publicly available as the "gap" program from Genetics Computer Group, Madison WI. The aforementioned parameters are the default parameters for polynucleotide comparisons.

By way of example, a polynucleotide sequence of the present invention may be identical to the reference sequences of SEQ ID NO: 1 or SEQ ID NO:7, that is be 100% identical, or it may include up to a certain integer number of nucleotide alterations as compared to the reference sequence. Such alterations are selected from the group consisting of at least one nucleotide deletion, substitution, including transition and transversion, or insertion, and wherein said alterations may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among the nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence. The number of nucleotide alterations is determined by multiplying the total number of nucleotides in SEQ ID NO: 1 or SEQ ID NO:7 by the numerical percent of the respective percent identity (divided by 100) and subtracting that product from said total number of nucleotides in SEQ ID NO: 1 or SEQ ID NO:7, or:

$$n_n \leq x_n - (x_n \bullet y),$$

wherein $\mathbf{n_n}$ is the number of nucleotide alterations, $\mathbf{x_n}$ is the total number of nucleotides in SEQ ID NO:1 or SEQ ID NO:7, and y is, for instance, 0.70 for 70%, 0.80 for 80%, 0.85 for 85%, 0.90 for 90%, 0.95 for 95%,etc., and wherein any non-integer product of $\mathbf{x_n}$ and $\mathbf{y}$ is rounded down to the nearest integer prior to subtracting it from $\mathbf{x_n}$. Alterations of a polynucleotide sequence encoding the polypeptide of SEQ ID NO:2 or SEQ ID NO:8 may create nonsense, missense or frameshift mutations in this coding sequence and thereby alter the polypeptide encoded by the polynucleotide following such alterations.

Similarly, a polypeptide sequence of the present invention may be identical to the reference sequences of SEQ ID NO:2 or SEQ ID NO:8, that is be 100% identical, or it may include up to a certain integer number of amino acid alterations as compared to the reference sequence such that the % identity is less than 100%. Such alterations are selected from the group consisting of at least one amino acid deletion, substitution, including conservative and non-conservative substitution, or insertion, and wherein said alterations may occur at the amino- or carboxy-terminal positions of the reference polypeptide sequence or anywhere between those terminal positions, interspersed either individually among the amino acids in the reference sequence or in one or more contiguous groups within the reference sequence. The number of amino acid alterations for a given % identity is determined by multiplying the total number of amino acids in SEQ ID NO:2 or SEQ ID NO:8 by the numerical percent of the respective percent identity(divided by 100) and then subtracting that product from said total number of amino acids in SEQ ID NO:2 or SEQ ID NO:8, or:

$$n_a \le x_a - (x_a \bullet y),$$

wherein $\mathbf{n_a}$ is the number of amino acid alterations, $\mathbf{x_a}$ is the total number of amino acids in SEQ ID NO:2 or SEQ ID NO:8, and $\mathbf{y}$ is, for instance 0.70 for 70%, 0.80 for 80%, 0.85 for 85% etc., and wherein any non-integer product of $\mathbf{x_a}$ and $\mathbf{y}$ is rounded down to the nearest integer prior to subtracting it from $\mathbf{x_a}$.

"Homolog" is a generic term used in the art to indicate a polynucleotide or polypeptide sequence possessing a high degree of sequence relatedness to a subject sequence. Such relatedness may be quantified by determining the degree of identity and/or similarity between the sequences being compared as hereinbefore described.

Falling within this generic term are the terms "ortholog", meaning a polynucleotide or polypeptide that is the functional equivalent of a polynucleotide or polypeptide in another species, and "paralog" meaning a functionally similar sequence when considered within the same species. Hence in the rat, for example, a member of the family of serotonin receptors is a paralog of the other members of the rat serotonin receptor family.

"Fusion protein" refers to a protein encoded by two, often unrelated, fused genes or fragments thereof. In one example, EP-A-0 464 discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, employing an immunoglobulin Fc region as a part of a fusion protein is advantageous for use in therapy and diagnosis resulting in, for example, improved pharmacokinetic properties [see, e.g., EP-A 0232 262]. On the other hand, for some uses it would be desirable to be able to delete the Fc part after the fusion protein has been expressed, detected and purified.

All publications, including but not limited to patents and patent applications, cited in this specification are herein incorporated by reference as if each individual publication were specifically and individually indicated to be incorporated by reference herein as though fully set forth.

SEQUENCE INFORMATION

SEQ ID NO:1

CACGCAGGGG CTGACAGCTG TGCTGGTGCT GATAAGGGAA GCCACAAGGA GACGATCGAG
GAGAGAGACA AGCGGCAGCA GAGGCAGCAG CGGCAGAGGC AGCACCAGGG CTGCGGAGCT
GCTGGGAGTG GGAGTGACTC CCCCACCTCG GGCCCCCACC CTGTCCCTGT CCTCTTCCCG
CTTGCCCTGA GTTTAGAAAA GCAGCCGCTG CCACCACTGC CACTCGGGAC GGCACCAGGC
CTGCTGGCTA GGGAGGGACA GGGCAGGGAG GCTCTGGCCA GTCCCAGCAG CCGGGGACAG
ATGCCGATCG AGATTGTGTG CAAAATCAAA TTTGCTGAGG AGGATGCGAA ACCCAAGGAG
AAGGAGGCAG GGGATGAGCA GAGCCTCCTC GGGGCTGTTG CCCCTGGAGC AGCGCCCCGA
GACCTGGCCA CCTTTGCCAG CACCAGCACC CTGCATGGAC TGGGCCGGGC CTGTGGCCCA
GGCCCCCACG GACTGCGCAG AACCCTGTGG GCACTGGCCC TACTCACCTC GCTGGCTGCC
TTCCTGTACC AGGCGGCTGG CCCGGCCCGG GGCTACCTGA CCCGGCCTCA CCTGGTGGCA
ATGGACCCCG CTGCCCCAGC CCCAGTGGCG GGCTTCCCGG CTGTCACCCT CTGCAATATC
AACCGCTTCC GGCATTCGGC ACTCAGCGAT GCCGACATCT TCCACCTGGC CAATCTGACA
GGGCTGCCCC CCAAAGACCG GGATGGGCAC CGTGCGGCTG GCCTGCGCTA CCCAGAGCCT
GACATGGTAG ACATCCTCAA CCGCACTGGC CACCAGCTCG CCGACATGCT TAAGAGCTGC
AACTTCAGTG GGCATCACTG CTCCGCCAGC AACTTCTCTG TGGTCTATAC TCGCTATGGG
AAGTGTTACA CCTTCAACGC GGACCCGCGG AGCTCCCTGC CCAGCCGGGC AGGGGGCATG
GGCAGTGGCC TGGAGATCAT GCTGGACATC CAGCAGGAGG AGTACCTGCC CATCTGGAGG
GAGACAAATG AGACGTCGTT TGAGGCAGGT ATTCGGGTGC AGATCCACAG CCAGGAGGAG
CCGCCCTACA TCCACCAGCT GGGGTTCGGG GTGTCCCCAG GCTTCCAGAC CTTTGTGTCC
TGCCAGGAAC AGCGGCTGAC CTACCTGCCC CAGCCCTGGG GCAACTGCCG CGCAGAGAGT
GAGCTCAGGG AGCCTGAGCT TCAGGGCTAC TCGGCCTACA GTGTGTCTGC CTGCCGGCTG
CGCTGTGAAA AGGAGGCCGT GCTTCAGCGC TGCCACTGCC GGATGGTGCA CATGCCAGGC
AATGAGACCA TCTGCCCACC AAATATCTAC ATCGAGTGTG CAGACCACAC ACTGGACTCC
CTGGGTGGGG GCCCTGAGGG CCCGTGCTTC TGCCCCACCC CCTGCAACCT GACGCGCTAT
GGGAAAGAGA TCTCCATGGT CAGGATCCCC AACAGGGGCT CAGCCCGGTA CCTGGCGAGG
AAGTACAACC GCAACGAGAC CTACATACGG GAGAACTTCC TGGTCCTAGA TGTCTTCTTT
GAGGCCCTGA CCTCTGAAGC CATGGAGCAG CGAGCAGCCT ATGGCCTGTC AGCCCTGCTG
GGAGACCTCG GGGGACAGAT GGGCCTGTTC ATTGGGGCCA GCATCCTCAC GTTGCTGGAG
ATCCTCGACT ACATCTATGA GGTGTCCTGG GATCGACTGA AGCGGGTATG GAGGCGTCCC
AAGACCCCCC TGCGGACCTC CACTGGGGGC ATCTCCACTT TGGGGCTTCA GGAGCTGAAG
GAACAGAGTC CCTGCCAGAG CCGGGGCCGA GTGGAGGGTG GGGGGGTCAG CAGTCTGCTC
CCCAATCACC ACCACCCCCA CGGTCCCCCA GGAGGTCTCT TTGAAGATTT TGCTTGCTAG
GACGGTGCTG TGACTGAAAG GACCCAGAGT CTGGGACCCC TCCTGGGATC CCCACACATT
CTCCTGCTCC TGGGAAAAAG CCTGGGGGCG GTGCTCACTC GCAAGGCCAA GAACTCAGTT
CCTGCTCTCA TCCTCCCCTG CCCTGATGTC ACTGCTTTGC ACAAAGGTCC TTCTTGTCCA
CACCCCTTAT CCCCAAGGCT GGTGCCCCGG GAGGGCTGGA GAACCAGGCC ATGGGCCCTC
ACGGAGAGGA AGGGAAGGAA GGAGAGGGAG GGGGAGGATA GAGCCCATCC CAGCCGGGGA
GGGGGAGCCT TCTGTACATT TGTAAATATT TAGGGAAAGC CGGGTGGGGG GAGGGGATAC
AGATGTAGAA GGTGGGTAGC GCTAACAGGG GTGGCTGATT TAGGGACAGC CAGGTCCCAG
CCCCAATGTC AGCAGGATAG GGAGAGCCCC AGGATTCAGG AGTGCTGGGC TGGTCCTACT
TCCTGCCCCT CTCCAGGCCC AGCTCCCTTT TTGGCAGGGG GAGAGGATGG CCCAGCAGGC
CTGGCCCAGT TCCCAGTTCC CCCTGCACCA GCCCCACCCC TAGAGTCCCT TTTATAGGGA
GGGGGCAGGA GACCTTCCAG ACTTCGGCTG AGCTTGGAGG GTGGGAAGGG AGCCTTCTCA
GTCCTCTCTC CCTCCAGTCT GATTTTATAA AGTGCTGACG AGATTGGGAA TAAAGAGGCA
TAAAGAATAA AAAAAAAAAA AAAAGAACTT GAGGGGGCCC TNTACAAAAG GCCCCTATAT
GATAGATANA T

SEQ ID NO:2

MPIEIVCKIK FAEEDAKPKE KEAGDEQSLL GAVAPGAAPR DLATFASTST LHGLGRACGP
GPHGLRRTLW ALALLTSLAA FLYQAAGPAR GYLTRPHLVA MDPAAPAPVA GFPAVTLCNI

NRFRHSALSD ADIFHLANLT GLPPKDRDGH RAAGLRYPEP DMVDILNRTG HQLADMLKSC
NFSGHHCSAS NFSVVYTRYG KCYTFNADPR SSLPSRAGGM GSGLEIMLDI QQEEYLPIWR
ETNETSFEAG IRVQIHSQEE PPYIHQLGFG VSPGFQTFVS CQEQRLTYLP QPWGNCRAES
ELREPELQGY SAYSVSACRL RCEKEAVLQR CHCRMVHMPG NETICPPNIY IECADHTLDS
LGGGPEGPCF CPTPCNLTRY GKEISMVRIP NRGSARYLAR KYNRNETYIR ENFLVLDVFF
EALTSEAMEQ RAAYGLSALL GDLGGQMGLF IGASILTLLE ILDYIYEVSW DRLKRVWRRP
KTPLRTSTGG ISTLGLQELK EQSPCQSRGR VEGGGVSSLL PNHHHPHGPP GGLFEDFAC

SEQ ID NO:3

TTGCCAACTA NATGCATGCT CGACCGGCCG CCAGTGTGAT GGATATCTGC AGAATTCGCC
CTTACTCACT ATAGGGCTCG AGCGGCCGCC CGGGCAGGTC ACGCGGGGCT GACAGCTGTG
CTGGTGCTGA TAAGGGAAGC CGGCCGCGGG ATTACTCACT ATAGGGCTCG AGCGGCCGCC
CGGGCAGGTG GCAGAGGCAG CACCAGGGCT GCGGAGCTGC TGGGAGTGGG AGTGACTCCC
CCACCTCGGG CCCCCACCCT GTCCCTGTCC TCTTCCCGCT TGCCCTGAGT TTAKAARAGC
AGCCGCTGCC ACCACTGCCA CTCGGGAGGG CACCAGGGCT GCTGGCTAGG GACGGACAGG
GCAGGGAGGC TCTGGCCAGT CCCAGCAGCC GGGGACAGAT GCCGATCGAG ATTGTGTGCA
AAATCAAATT TGCTGAGGAG GATGCGAAAC CCAAGGAGAA GGAGGCAGGG GATGAGCAGA
GCCTCCTCGG GGCTGTTGCC CCTGGAGCAG CGCCCCGAGA CCTGGCCACC TTTGCCAGCA
CCAGCACCCT GCATGGACTG GGCCGGGCCT GTGGCCCAGG CCCCCACGGA CTGCGCAGAA
CCCTGTGGGC ACTGGCCCTA CTCACCTCGC TGGCTGCCTT CCTGTACCAG GCGGCTGGCC
TGGCCCGGGG CTACCTGACC CGGCCTCACC TGGTGGCAAT GGACCCCGCT GCCCCAGCCC
CAGTGGCGGG CTTCCCGGCT GTCACCCTCT GCAATATCAA CCGCTTCCGG CATTCGGCAC
TCAGCGATGC CGACATCTTC CACCTGGCCA ATCTGACAGG GCTGCCCCCC AAAGACCGGG
ATGGGCACCG TGCGGCTGGC CTGCGCTACC CAGAGCCTGA CATGGTAGAC ATCCTCAACC
GCACTGGCCA CCAGCTCGCC GACATGCTTA AGAGCTGCAA CTTCAGTGGG CATCACTGCT
CCGCCAGCAA CTTCTCTGTG GTCTATACTC GCTATGGGAA GTGTTACACC TTCAACGCGG
ACCCGCGGAG CTCGCTGCCC AGCCGGGCAG GGGGCATGGG CAGTGGCCTG GAAATCATGC
TGGACATCCA GCAGGAGGAG TACCTGCCCA TCTGGAGGGA GACAAATGAG ACGTCGTTTG
AGGCAGGTAT TCGGGTGCAG ATCCACAGCC AGGACGAGCC GCCCTACATC CACCAGCTGG
GGTTCGGGGT GTCCCCAGGC TTCCAGACCT TTGTGTCCTG CCAGGAACAG CGGCTGACCT
ACCTGCCCCA GCCCTGGGGC AACTGCCGCG CAGAGAGTGA GCTCAGGGAG CCTGAGCTTC
AGGGCTACTC GGCCTACAGT GTGTCTGCCT GCCGGCTGCG CTGTGAAAAG GAGGCCGTGY
TTCAGCGCTG CCACTGCCGG ATGGTGCACA TGCCAGGCAA TGAGACCATY TGCCCACCAA
ATATTTACAT CGAGTGTGCA GACCACACAC TGGGAAAAAT GCATTGCTGG CTATCGGTGT
GGGTAAGGCA AACCCCATCT GTCCACAGAT CTTGCCTGAT CTGCAGTTCA TCCATCTATC
CATCCATTCA TCGATTCGCC CATTCTCCAG ACCTTTACAG CCTGTGCTGG GTACTGGAGA
CTCCCTGGGT GGGGGCCCTG AGGGCCCGTG CTTTTGCCCC ACCCCCTGCA ACCTGACACG
CTATGGGAAA GAGATCTCCA TGGTCAGGAT CCCCAACAGG GGCTCAGCCC GGTACCTGAC
GAGGAAGTAC AACCGCAACG AGACCTACAT ACGGGAGAAC TTCCTGGTCC TAGATGTCTT
CTTTGAGGCC CTGACCTCTG AAGCCATGGA GCAGCGAGCA GCCTATGGCC TGTCAGCCCT
GCTGGGAGAC CTCGGGGGAC AGATGGGCCT GTTCATTGGG GCCAGCATCC TCACGTTGCT
GGAGATCCTC GACTACATCT ATGARGTGTC CTGGGATCGA CTGAAGCGGG TATGGAGGCG
TCCCAAGACC CCCTGCGGAC CTCCACTGGG GGCATCTCCA CTTTGGGGCT TCAGGAGCTG
AAGGAACARA TCCCTGCCCR ACCGGGGCCG AWTGGAAGGT GGGGGGGTCA RCARTCTGCT

CCCAATCACC ACCACCCCCA CGGTCCCCCA GGAAGTCTCT TTGAAAATTT TGCTTGCTAG
GACGGTGCTG TACTGAAAGG ACCCAGAATT CTGGGACCCC TCCTGGGATC CCCANCACAT
TCTCCTGCTC CTGGGAAAAA GCCTGGCGGC GGTGCTCACT GGGAAGGCCA AGAACTCAGT
TCCTGCTCTC ATCCTCCCCT GCCCTGATGT CACTGCTTTG CACAAAGGTC CTTCTTGTCC
ACACCCCYTT AWTCCCCAAG GCTGGTGCCC CGGRAGGGCT GGAGAACCAG GCCATGGGCC
CTTCACGGAG AGGAAGGGAA GGAAGGAGAG GGAGGGGGAG GATAGAGCCC ATCCCAGCCG
GGGAGGGGGA GCCTTCTGTA CATTTGTAAA TATTTAGGGA AAGCCGGGTG GGGGGAGGGG
ATACAGATGT AGAAGGTGGG TAGGGCTAAC AGGGGTCGGT GATTTAGGGA CAGCCAGGTC
CCAGCCCCAA TGTCAGCAGG ATAGGGAGAG CCCCAGGATT CAGGAGTGCT GGGCTGGTCC
TACTTCCTGC CCCTCTCCAG GCCCAGCTCC CTTTTTGGCA GGGGGAGAGG ATGGCCCAGC
AGGCCTGGCC CAGTTCCCAG TTCCCCCTGC ACCAGCCCCA CCCCTAGAGT CCCTTTTATA
GGGAGGGGGC AGGAGACCTT CCAGACTTCG GCTGAGCTTG GAGGGTGGGA AGGGAGCCTT
CTCAGTCCTC TCTCCCTCCA GTCTGATTTT ATAAAGTGCT GACGAGATTG GGAATAAAGA
GGCATAAAGA ATAAAAAAAA AAAAAAAAGA ACTTGAGGGG GGCCTNTACA AAAGGCCCCT
ATATGATAGA TANAT

SEQ ID NO:4

MPIEIVCKIK FAEEDAKPKE KEAGDEQSLL GAVAPGAAPR DLATFASTST LHGLGRACGP
GPHGLRRTLW ALALLTSLAA FLYQAAGLAR GYLTRPHLVA MDPAAPAPVA GFPAVTLCNI
NRFRHSALSD ADIFHLANLT GLPPKDRDGH RAAGLRYPEP DMVDILNRTG HQLADMLKSC
NFSGHHCSAS NFSVVYTRYG KCYTFNADPR SSLPSRAGGM GSGLEIMLDI QQEEYLPIWR
ETNETSFEAG IRVQIHSQEE PPYIHQLGFG VSPGFQTFVS CQEQRLTYLP QPWGNCRAES
ELREPELQGY SAYSVSACRL RCEKEAVXQR CHCRMVHMPG NETICPPNIY IECADXXXXX
XXXXXXXXXX XXXXXXXXXX XXXXXXXXXX XXXXXXXXXX XXXXXXXXXX XXXXXXXCFC
PTPCNLTRYG KEISMVRIPN RGSARYLTRK YNRNETYIRE NFLVLDVFFE ALTSEAMEQR
AAYGLSALLG DLGGQMGLFI GASILTLLEI LDYIYEVSWD RLKRXMEASQ DPLRTSTGGI
STLGLQELKE QIPAXPGPXG RWGGQQSAPN HHHPHGPPGS LFENFAC

SEQ ID NO:5

AATTCGGCAC GAGTGCAACC TGACACGCTA TGGGNAAGAG ATCTCCATGG TCAGGATCCC
CAACAGGGGC TCAGCCCGGT ACCTGACGAG GAAGTACAAC CGCAACGAGA CCTACATACG
GGAGAACTTC CTGGTACTAG ATGTCTTCTT TGAGGCCCTG ACCTNTGAAG NCATGGAGCA
GCGAGCAGCC TATGGCCTGT CAGCCCTGCT GGGAGACCTC GGGGGACAGA TGGGNCTGTT
CATTGGGGCC AGCATCCTCA CGTTGCTGGA GATTCTTGAC TACATCTATG AGGTGTCCTG
GGGATCGACT TGAAGCGGGT ATTGGAGGCG TNCCNAGACC CCCCTTGNGG GACCTNCACT
TGGGGGGATT TCCAATTTTG GGGNTTCAAG AGGTTTNAGG AACANNNTTC CTNCCCNAGC
CGGGGCCCAT TTGAGGGTTG GGGGGTCAAA ANTTTTTTCC ATTAACACCA CCCCAGGTCC
CAGNAGTTTT TNAGGTTTTT TTTAGGNG

SEQ ID NO:6

IRHECNLTRY GXEISMVRIP NRGSARYLTR KYNRNETYIR ENFLVLDVFF EALTXEXMEQ
RAAYGLSALL GDLGGQMGLF IGASILTLLE ILDYIYEVSW GST

SEQ ID NO:7

ATGCCGATCG AGATTGTGTG CAAAATCAAA TTTGCTGAGG AGGATGCGAA ACCCAAGGAG
AAGGAGGCAG GGGATGAGCA GAGCCTCCTC GGGGCTGTCG CCCCTGGAGC AGCCCCCCGA
GACCTGGCCA CCTTTGCCAG CACCAGCACC CTGCATGGAC TGGGCCGGGC CTGTGGCCCA
GGCCCCCACG GACTGCGCAG AACCCTGTGG GCACTGGCCC TACTCACCTC GCTGGCTGCC
TTCCTGTACC AGGCGGCTGG CCTGGCCCGG GGCTACCTGA CCCGGCCTCA CCTGGTGGCA
ATGGACCCCG CTGCCCCAGC CCCAGTGGCG GGCTTCCCGG CTGTCACCCT CTGCAATATC
AACCGCTTCC GGCATTCGGC ACTCAGCGAT GCCGACATCT TCCACCTGGC CAATCTGACA
GGGCTGCCCC CCAAAGACCG GGATGGGCAC CGTGCGGCTG GCCTGCGCTA CCCAGAGCCT
GACATGGTAG ACATCCTCAA CCGCACTGGC CACCAGCTCG CCGACATGCT TAAGAGCTGC
AACTTCAGTG GGCATCACTG CTCCGCCAGC AACTTCTCTG TGGTCTATAC TCGCTATGGG
AAGTGTTACA CCTTCAACGC GGACCCGCGG AGCTCGCTGC CCAGCCGGGC AGGGGGCATG
GGCAGTGGCC TGGAGATCAT GCTGGACATC CAGCAGGAGG AGTACCTGCC CATCTGGAGG
GAGACAAATG AGACGTCGTT TGAGGCAGGT ATTCGGGTGC AGATCCACAG CCAGGAGGAG
CCGCCCTACA TCCACCAGCT GGGGTTCGGG GTGTCCCCAG GCTTCCAGAC CTTTGTGTCC
TGCCAGGAAC AGCGGCTGAC CTACCTGCCC CAGCCCTGGG GCAACTGCCG CGCAGAGAGT
GAGCTCAGGG AGCCTGAGCT TCAGGGCTAC TCGGCCTACA GTGTGTCTGC CTGCCGGCTG
CGCTGTGAAA AGGAGGCCGT GCTTCAGCGC TGCCACTGCC GGATGGTGCA CATGCCAGGC
AATGAGACCA TCTGCCCACC AAATATCTAC ATCGAGTGTG CAGACCACAC ACTGGACTCC
CTGGGTGGGG GCCCTGAGGG CCCGTGCTTC TGCCCCACCC CCTGCAACCT GACACGCTAT
GGGAAAGAGA TCTCCATGGT CAGGATCCCC AACAGGGGCT CAGCCCGGTA CCTGGCGAGG
AAGTACAACC GCAACGAGAC CTACATACGG GAGAACTTCC TGGTCCTAGA TGTCTTCTTT
GAGGCCCTGA CCTCTGAAGC CATGGAGCAG CGAGCAGCCT ATGGCCTGTC AGCCCTGCTG
GGAGACCTCG GGGGACAGAT GGGCCTGTTC ATTGGGGCCA GCATCCTCAC GTTGCTGGAG
ATCCTCGACT ACATCTATGA GGTGTCCTGG GATCGACTGA AGCGGGTATG GAGGCGTCCC
AAGACCCCCC TGCGGACCTC CACTGGGGGC ATCTCCACTT TGGGGCTTCA GGAGCTGAAG
GAACAGAGTC CCTGTCCGAG CCTGGGCCGA GCGGAGGGTG GGGGGGTCAG CAGTCTGCTC
CCCAATCACC ACCACCCCCA CGGTCCCCCA GGAGGTCTCT TTGAAGATTT TGCTTGCTAG
GACGGTGCTG TG

SEQ ID NO:8

MPIEIVCKIK FAEEDAKPKE KEAGDEQSLL GAVAPGAAPR DLATFASTST LHGLGRACGP
GPHGLRRTLW ALALLTSLAA FLYQAAGLAR GYLTRPHLVA MDPAAPAPVA GFPAVTLCNI
NRFRHSALSD ADIFHLANLT GLPPKDRDGH RAAGLRYPEP DMVDILNRTG HQLADMLKSC
NFSGHHCSAS NFSVVYTRYG KCYTFNADPR SSLPSRAGGM GSGLEIMLDI QQEEYLPIWR
ETNETSFEAG IRVQIHSQEE PPYIHQLGFG VSPGFQTFVS CQEQRLTYLP QPWGNCRAES
ELREPELQGY SAYSVSACRL RCEKEAVLQR CHCRMVHMPG NETICPPNIY IECADHTLDS
LGGGPEGPCF CPTPCNLTRY GKEISMVRIP NRGSARYLAR KYNRNETYIR ENFLVLDVFF
EALTSEAMEQ RAAYGLSALL GDLGGQMGLF IGASILTLLE ILDYIYEVSW DRLKRVWRRP
KTPLRTSTGG ISTLGLQELK EQSPCPSLGR AEGGGVSSLL PNHHHPHGPP GGLFEDFAC

Annex to the description

SEQUENCE LISTING

(1) GENERAL INFORMATION

(I) APPLICANT: SMITHKLINE BEECHAM PLC

(II) TITLE OF THE INVENTION: NOVEL COMPOUNDS

(III) NUMBER OF SEQUENCES: 8

(IV) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: SMITHKLINE BEECHAM
(B) STREET: NEW HORIZONS COURT, GREAT WEST ROAD
(C) CITY: BRENTFORD
(D) STATE:
(E) COUNTRY: UK
(F) ZIP: TW8 9ET

(V) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: DISKETTE
(B) COMPUTER: IBM COMPATIBLE
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FASTSEQ FOR WINDOWS VERSION 2.0

(VI) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(VII) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:

(VIII) ATTORNEY/AGENT INFORMATION:
(A) NAME: CONNELL, CHRIS
(B) REGISTRATION NUMBER:
(C) REFERENCE/DOCKET NUMBER: GH30021

(IX) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 0181-9752000

(B) TELEFAX: 0181-9756294

(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(I) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2711 BASE PAIRS

(B) TYPE: NUCLEIC ACID

(C) STRANDEDNESS: SINGLE

(D) TOPOLOGY: LINEAR

(II) MOLECULE TYPE: CDNA

(XI) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
CACGCAGGGG CTGACAGCTG TGCTGGTGCT GATAAGGGAA GCCACAAGGA GACGATCGAG    60
GAGAGAGACA AGCGGCAGCA GAGGCAGCAG CGGCAGAGGC AGCACCAGGG CTGCGGAGCT   120
GCTGGGAGTG GGAGTGACTC CCCCACCTCG GGCCCCCACC CTGTCCCTGT CCTCTTCCCG   180
CTTGCCCTGA GTTTAGAAAA GCAGCCGCTG CCACCACTGC CACTCGGGAG GGCACCAGGG   240
CTGCTGGCTA GGGAGGGACA GGGCAGGGAG GCTCTGGCCA GTCCCAGCAG CCGGGGACAG   300
ATGCCGATCG AGATTGTGTG CAAAATCAAA TTTGCTGAGG AGGATGCGAA ACCCAAGGAG   360
AAGGAGGCAG GGGATGAGCA GAGCCTCCTC GGGGCTGTTG CCCCTGGAGC AGCGCCCCGA   420
GACCTGGCCA CCTTTGCCAG CACCAGCACC CTGCATGGAC TGGGCCGGGC CTGTGGCCCA   480
GGCCCCCACG GACTGCGCAG AACCCTGTGG GCACTGGCCC TACTCACCTC GCTGGCTGCC   540
TTCCTGTACC AGGCGGCTGG CCCGGCCCGG GGCTACCTGA CCCGGCCTCA CCTGGTGGCA   600
ATGGACCCCG CTGCCCCAGC CCCAGTGGCG GGCTTCCCGG CTGTCACCCT CTGCAATATC   660
AACCGCTTCC GGCATTCGGC ACTCAGCGAT GCCGACATCT TCCACCTGGC CAATCTGACA   720
GGGCTGCCCC CCAAAGACCG GGATGGGCAC CGTGCGGCTG GCCTGCGCTA CCCAGAGCCT   780
GACATGGTAG ACATCCTCAA CCGCACTGGC CACCAGCTCG CCGACATGCT TAAGAGCTGC   840
AACTTCAGTG GGCATCACTG CTCCGCCAGC AACTTCTCTG TGGTCTATAC TCGCTATGGG   900
AAGTGTTACA CCTTCAACGC GGACCCGCGG AGCTCGCTGC CCAGCCGGGC AGGGGGCATG   960
GGCAGTGGCC TGGAGATCAT GCTGGACATC CAGCAGGAGG AGTACCTGCC CATCTGGAGG  1020
GAGACAAATG AGACGTCGTT TGAGGCAGGT ATTCGGGTGC AGATCCACAG CCAGGAGGAG  1080
CCGCCCTACA TCCACCAGCT GGGGTTCGGG GTGTCCCCAG GCTTCCAGAC CTTTGTGTCC  1140
TGCCAGGAAC AGCGGCTGAC CTACCTGCCC CAGCCCTGGG GCAACTGCCG CGCAGAGAGT  1200
GAGCTCAGGG AGCCTGAGCT TCAGGGCTAC TCGGCCTACA GTGTGTCTGC CTGCCGGCTG  1260
CGCTGTGAAA AGGAGGCCGT GCTTCAGCGC TGCCACTGCC GGATGGTGCA CATGCCAGGC  1320
AATGAGACCA TCTGCCCACC AAATATCTAC ATCGAGTGTG CAGACCACAC ACTGGACTCC  1380
CTGGGTGGGG GCCCTGAGGG CCCGTGCTTC TGCCCCACCC CCTGCAACCT GACGCGCTAT  1440
GGGAAAGAGA TCTCCATGGT CAGGATCCCC AACAGGGGCT CAGCCCGGTA CCTGGCGAGG  1500
AAGTACAACC GCAACGAGAC CTACATACGG GAGAACTTCC TGGTCCTAGA TGTCTTCTTT  1560
GAGGCCCTGA CCTCTGAAGC CATGGAGCAG CGAGCAGCCT ATGGCCTGTC AGCCCTGCTG  1620
GGAGACCTCG GGGGACAGAT GGGCCTGTTC ATTGGGGCCA GCATCCTCAC GTTGCTGGAG  1680
ATCCTCGACT ACATCTATGA GGTGTCCTGG GATCGACTGA AGCGGGTATG GAGGCGTCCC  1740
AAGACCCCCC TGCGGACCTC CACTGGGGGC ATCTCCACTT TGGGGCTTCA GGAGCTGAAG  1800
```

```
GAACAGAGTC CCTGCCAGAG CCGGGGCCGA GTGGAGGGTG GGGGGGTCAG CAGTCTGCTC  1860
CCCAATCACC ACCACCCCCA CGGTCCCCCA GGAGGTCTCT TTGAAGATTT TGCTTGCTAG  1920
GACGGTGCTG TGACTGAAAG GACCCAGAGT CTGGGACCCC TCCTGGGATC CCCACACATT  1980
CTCCTGCTCC TGGGAAAAAG CCTGGGGGCG GTGCTCACTG GGAAGGCCAA GAACTCAGTT  2040
CCTGCTCTCA TCCTCCCCTG CCCTGATGTC ACTGCTTTGC ACAAAGGTCC TTCTTGTCCA  2100
CACCCCTTAT CCCCAAGGCT GGTGCCCCGG GAGGGCTGGA GAACCAGGCC ATGCGCCCTC  2160
ACGGAGAGGA AGGGAAGGAA GGAGAGGGAG GGGGAGGATA GAGCCCATCC CAGCCGGGGA  2220
GGGGGAGCCT TCTGTACATT TGTAAATATT TAGGGAAAGC CGGGTGGGGG GAGGGGATAC  2280
AGATGTAGAA GGTGGGTAGG GCTAACAGGG GTGGGTGATT TAGGGACAGC CAGGTCCCAG  2340
CCCCAATGTC AGCAGGATAG GGAGAGCCCC AGGATTCAGG AGTGCTGGGC TGGTCCTACT  2400
TCCTGCCCCT CTCCAGGCCC AGCTCCCTTT TTGGCAGGGG GAGAGGATGG CCCAGCAGGC  2460
CTGGCCCAGT TCCCAGTTCC CCCTGCACCA GCCCCACCCC TAGAGTCCCT TTTATAGGGA  2520
GGGGGCAGGA GACCTTCCAG ACTTCGGCTG AGCTTGGAGG GTGGGAAGGG AGCCTTCTCA  2580
GTCCTCTCTC CCTCCAGTCT GATTTTATAA AGTGCTGACG AGATTGGGAA TAAAGAGGCA  2640
TAAAGAATAA AAAAAAAAAA AAAAGAACTT GAGCGGGGCC TNTACAAAAG GCCCCTATAT  2700
GATAGATANA T                                                       2711
```

(2) INFORMATION FOR SEQ ID NO:2:

(I) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 539 AMINO ACIDS

(B) TYPE: AMINO ACID

(C) STRANDEDNESS: SINGLE

(D) TOPOLOGY: LINEAR

(II) MOLECULE TYPE: PROTEIN

(XI) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
MET PRO ILE GLU ILE VAL CYS LYS ILE LYS PHE ALA GLU GLU ASP ALA
 1               5                  10                  15
LYS PRO LYS GLU LYS GLU ALA GLY ASP GLU GLN SER LEU LEU GLY ALA
            20                  25                  30
VAL ALA PRO GLY ALA ALA PRO ARG ASP LEU ALA THR PHE ALA SER THR
        35                  40                  45
SER THR LEU HIS GLY LEU GLY ARG ALA CYS GLY PRO GLY PRO HIS GLY
    50                  55                  60
LEU ARG ARG THR LEU TRP ALA LEU ALA LEU LEU THR SER LEU ALA ALA
65                  70                  75                  80
PHE LEU TYR GLN ALA ALA GLY PRO ALA ARG GLY TYR LEU THR ARG PRO
                85                  90                  95
HIS LEU VAL ALA MET ASP PRO ALA ALA PRO ALA PRO VAL ALA GLY PHE
            100                 105                 110
PRO ALA VAL THR LEU CYS ASN ILE ASN ARG PHE ARG HIS SER ALA LEU
        115                 120                 125
SER ASP ALA ASP ILE PHE HIS LEU ALA ASN LEU THR GLY LEU PRO PRO
```

```
    130                 135                 140
LYS ASP ARG ASP GLY HIS ARG ALA ALA GLY LEU ARG TYR PRO GLU PRO
145                 150                 155                 160
ASP MET VAL ASP ILE LEU ASN ARG THR GLY HIS GLN LEU ALA ASP MET
                165                 170                 175
LEU LYS SER CYS ASN PHE SER GLY HIS HIS CYS SER ALA SER ASN PHE
            180                 185                 190
SER VAL VAL TYR THR ARG TYR GLY LYS CYS TYR THR PHE ASN ALA ASP
        195                 200                 205
PRO ARG SER SER LEU PRO SER ARG ALA GLY GLY MET GLY SER GLY LEU
    210                 215                 220
GLU ILE MET LEU ASP ILE GLN GLN GLU GLU TYR LEU PRO ILE TRP ARG
225                 230                 235                 240
GLU THR ASN GLU THR SER PHE GLU ALA GLY ILE ARG VAL GLN ILE HIS
                245                 250                 255
SER GLN GLU GLU PRO PRO TYR ILE HIS GLN LEU GLY PHE GLY VAL SER
            260                 265                 270
PRO GLY PHE GLN THR PHE VAL SER CYS GLN GLU GLN ARG LEU THR TYR
        275                 280                 285
LEU PRO GLN PRO TRP GLY ASN CYS ARG ALA GLU SER GLU LEU ARG GLU
    290                 295                 300
PRO GLU LEU GLN GLY TYR SER ALA TYR SER VAL SER ALA CYS ARG LEU
305                 310                 315                 320
ARG CYS GLU LYS GLU ALA VAL LEU GLN ARG CYS HIS CYS ARG MET VAL
                325                 330                 335
HIS MET PRO GLY ASN GLU THR ILE CYS PRO PRO ASN ILE TYR ILE GLU
            340                 345                 350
CYS ALA ASP HIS THR LEU ASP SER LEU GLY GLY GLY PRO GLU GLY PRO
        355                 360                 365
CYS PHE CYS PRO THR PRO CYS ASN LEU THR ARG TYR GLY LYS GLU ILE
    370                 375                 380
SER MET VAL ARG ILE PRO ASN ARG GLY SER ALA ARG TYR LEU ALA ARG
385                 390                 395                 400
LYS TYR ASN ARG ASN GLU THR TYR ILE ARG GLU ASN PHE LEU VAL LEU
                405                 410                 415
ASP VAL PHE PHE GLU ALA LEU THR SER GLU ALA MET GLU GLN ARG ALA
            420                 425                 430
ALA TYR GLY LEU SER ALA LEU LEU GLY ASP LEU GLY GLY GLN MET GLY
        435                 440                 445
LEU PHE ILE GLY ALA SER ILE LEU THR LEU LEU GLU ILE LEU ASP TYR
    450                 455                 460
ILE TYR GLU VAL SER TRP ASP ARG LEU LYS ARG VAL TRP ARG ARG PRO
465                 470                 475                 480
LYS THR PRO LEU ARG THR SER THR GLY GLY ILE SER THR LEU GLY LEU
```

485 490 495
GLN GLU LEU LYS GLU GLN SER PRO CYS GLN SER ARG GLY ARG VAL GLU
500 505 510
GLY GLY GLY VAL SER SER LEU LEU PRO ASN HIS HIS HIS PRO HIS GLY
515 520 525
PRO PRO GLY GLY LEU PHE GLU ASP PHE ALA CYS
530 535

(2) INFORMATION FOR SEQ ID NO:3:

(I) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2955 BASE PAIRS
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(II) MOLECULE TYPE: CDNA

(XI) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
TTGCCAACTA NATGCATGCT CGACCGGCCG CCAGTGTGAT GGATATCTGC AGAATTCGCC     60
CTTACTCACT ATAGGGCTCG AGCGGCCGCC CGGGCAGGTC ACGCGGGGCT GACAGCTGTG    120
CTGGTGCTGA TAAGGGAAGC CGGCCGCGGG ATTACTCACT ATAGGGCTCG AGCGGCCGCC    180
CGGGCAGGTG GCAGAGGCAG CACCAGGGCT GCGGAGCTGC TGGGAGTGGG AGTGACTCCC    240
CCACCTCGGG CCCCCACCCT GTCCCTGTCC TCTTCCCGCT TGCCCTGAGT TTAKAARAGC    300
AGCCGCTGCC ACCACTGCCA CTCGGGAGGG CACCAGGGCT GCTGGCTAGG GAGGGACAGG    360
GCAGGGAGGC TCTGGCCAGT CCCAGCAGCC GGGGACAGAT GCCGATCGAG ATTGTGTGCA    420
AAATCAAATT TGCTGAGGAG GATGCGAAAC CCAAGGAGAA GGAGGCAGGG GATGAGCAGA    480
GCCTCCTCGG GGCTGTTGCC CCTGGAGCAG CGCCCCGAGA CCTGGCCACC TTTGCCAGCA    540
CCAGCACCCT GCATGGACTG GGCCGGGCCT GTGGCCCAGG CCCCCACGGA CTGCGCAGAA    600
CCCTGTGGGC ACTGGCCCTA CTCACCTCGC TGGCTGCCTT CCTGTACCAG GCGGCTGGCC    660
TGGCCCGGGG CTACCTGACC CGGCCTCACC TGGTGGCAAT GGACCCCGCT GCCCCAGCCC    720
CAGTGGCGGG CTTCCCGGCT GTCACCCTCT GCAATATCAA CCGCTTCCGG CATTCGGCAC    780
TCAGCGATGC CGACATCTTC CACCTGGCCA ATCTGACAGG GCTGCCCCCC AAAGACCGGG    840
ATGGGCACCG TGCGGCTGGC CTGCGCTACC CAGAGCCTGA CATGGTAGAC ATCCTCAACC    900
GCACTGGCCA CCAGCTCGCC GACATGCTTA AGAGCTGCAA CTTCAGTGGG CATCACTGCT    960
CCGCCAGCAA CTTCTCTGTG GTCTATACTC GCTATGGGAA GTGTTACACC TTCAACGCGG   1020
ACCCGCGGAG CTCGCTGCCC AGCCGGGCAG GGGGCATGGG CAGTGGCCTG GAAATCATGC   1080
TGGACATCCA GCAGGAGGAG TACCTGCCCA TCTGGAGGGA GACAAATGAG ACGTCGTTTG   1140
AGGCAGGTAT TCGGGTGCAG ATCCACAGCC AGGAGGAGCC GCCCTACATC CACCAGCTGG   1200
GGTTCGGGGT GTCCCCAGGC TTCCAGACCT TTGTGTCCTG CCAGGAACAG CGGCTGACCT   1260
ACCTGCCCCA GCCCTGGGGC AACTGCCGCG CAGAGAGTGA GCTCAGGGAG CCTGAGCTTC   1320
AGGGCTACTC GGCCTACAGT GTGTCTGCCT GCCGGCTGCG CTGTGAAAAG GAGGCCGTGY   1380
TTCAGCGCTG CCACTGCCGG ATGGTGCACA TGCCAGGCAA TGAGACCATY TGCCCACCAA   1440
ATATTTACAT CGAGTGTGCA GACCACACAC TGGGAAAAAT GCATTGCTGG CTATCGGTGT   1500
GGGTAAGGCA AACCCCATCT GTCCACAGAT CTTGCCTGAT CTGCAGTTCA TCCATCTATC   1560
CATCCATTCA TCGATTCGCG CATTCTCCAG ACCTTTACAG CCTGTGCTGG GTACTGGAGA   1620
```

```
CTCCCTGGGT GGGGGCCCTG AGGGCCCGTG CTTTTGCCCC ACCCCCTGCA ACCTGACACG   1680
CTATGGGAAA GAGATCTCCA TGGTCAGGAT CCCCAACAGG GGCTCAGCCC GGTACCTGAC   1740
GAGGAAGTAC AACCGCAACG AGACCTACAT ACGGGAGAAC TTCCTGGTCC TAGATGTCTT   1800
CTTTGAGGCC CTGACCTCTG AAGCCATGGA GCAGCGAGCA GCCTATGGCC TGTCAGCCCT   1860
GCTGGGAGAC CTCGGGGGAC AGATGGGCCT GTTCATTGGG GCCAGCATCC TCACGTTGCT   1920
GGAGATCCTC GACTACATCT ATGARGTGTC CTGGGATCGA CTGAAGCGGG TATGGAGGCG   1980
TCCCAAGACC CCCTGCGGAC CTCCACTGGG GGCATCTCCA CTTTGGGGCT TCAGGAGCTG   2040
AAGGAACARA TCCCTGCCCR ACCGGGGCCG AWTGGAAGGT GGGGGGGTCA RCARTCTGCT   2100
CCCAATCACC ACCACCCCCA CGGTCCCCCA GGAAGTCTCT TTGAAAATTT TGCTTGCTAG   2160
GACGGTGCTG TACTGAAAGG ACCCAGAATT CTGGGACCCC TCCTGGGATC CCCANCACAT   2220
TCTCCTGCTC CTGGGAAAAA GCCTGGGGGC GGTGCTCACT GGGAAGGCCA AGAACTCAGT   2280
TCCTGCTCTC ATCCTCCCCT GCCCTGATGT CACTGCTTTG CACAAAGGTC CTTCTTGTCC   2340
ACACCCCYTT AWTCCCCAAG GCTGGTGCCC CGGRAGGGCT GGAGAACCAG GCCATGGGCC   2400
CTTCACGGAG AGGAAGGGAA GGAAGGAGAG GGAGGGGGAG GATAGAGCCC ATCCCAGCCG   2460
GGGAGGGGGA GCCTTCTGTA CATTTGTAAA TATTTAGGGA AAGCCGGGTG GGGGGAGGGG   2520
ATACAGATGT AGAAGGTGGG TAGGGCTAAC ACGGGTGGGT GATTTAGGGA CAGCCAGGTC   2580
CCAGCCCCAA TGTCAGCAGG ATAGGGAGAG CCCCAGGATT CAGGAGTGCT GGGCTGGTCC   2640
TACTTCCTGC CCCTCTCCAG GCCCAGCTCC CTTTTTGGCA GGGGGAGAGG ATGGCCCAGC   2700
AGGCCTGGCC CAGTTCCCAG TTCCCCCTGC ACCAGCCCCA CCCCTAGAGT CCCTTTTATA   2760
GGGAGGGGGC AGGAGACCTT CCAGACTTCG GCTGAGCTTG GAGGGTGGGA AGGGAGCCTT   2820
CTCAGTCCTC TCTCCCTCCA GTCTGATTTT ATAAAGTGCT GACGAGATTG GGAATAAAGA   2880
GGCATAAAGA ATAAAAAAAA AAAAAAAAGA ACTTGAGGGG GGCCTNTACA AAAGGCCCCT   2940
ATATGATAGA TANAT                                                    2955
```

(2) INFORMATION FOR SEQ ID NO:4:

(I) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 587 AMINO ACIDS

(B) TYPE: AMINO ACID

(C) STRANDEDNESS: SINGLE

(D) TOPOLOGY: LINEAR

(II) MOLECULE TYPE: PROTEIN

(XI) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
MET PRO ILE GLU ILE VAL CYS LYS ILE LYS PHE ALA GLU GLU ASP ALA
 1               5                  10                  15
LYS PRO LYS GLU LYS GLU ALA GLY ASP GLU GLN SER LEU LEU GLY ALA
            20                  25                  30
VAL ALA PRO GLY ALA ALA PRO ARG ASP LEU ALA THR PHE ALA SER THR
        35                  40                  45
SER THR LEU HIS GLY LEU GLY ARG ALA CYS GLY PRO GLY PRO HIS GLY
    50                  55                  60
LEU ARG ARG THR LEU TRP ALA LEU ALA LEU LEU THR SER LEU ALA ALA
65                  70                  75                  80
PHE LEU TYR GLN ALA ALA GLY LEU ALA ARG GLY TYR LEU THR ARG PRO
```

```
                        85                      90                      95
HIS LEU VAL ALA MET ASP PRO ALA ALA PRO ALA PRO VAL ALA GLY PHE
                100                     105                     110
PRO ALA VAL THR LEU CYS ASN ILE ASN ARG PHE ARG HIS SER ALA LEU
            115                     120                     125
SER ASP ALA ASP ILE PHE HIS LEU ALA ASN LEU THR GLY LEU PRO PRO
        130                     135                     140
LYS ASP ARG ASP GLY HIS ARG ALA ALA GLY LEU ARG TYR PRO GLU PRO
145                     150                     155                     160
ASP MET VAL ASP ILE LEU ASN ARG THR GLY HIS GLN LEU ALA ASP MET
                        165                     170                     175
LEU LYS SER CYS ASN PHE SER GLY HIS HIS CYS SER ALA SER ASN PHE
                180                     185                     190
SER VAL VAL TYR THR ARG TYR GLY LYS CYS TYR THR PHE ASN ALA ASP
            195                     200                     205
PRO ARG SER SER LEU PRO SER ARG ALA GLY GLY MET GLY SER GLY LEU
        210                     215                     220
GLU ILE MET LEU ASP ILE GLN GLN GLU GLU TYR LEU PRO ILE TRP ARG
225                     230                     235                     240
GLU THR ASN GLU THR SER PHE GLU ALA GLY ILE ARG VAL GLN ILE HIS
                        245                     250                     255
SER GLN GLU GLU PRO PRO TYR ILE HIS GLN LEU GLY PHE GLY VAL SER
                260                     265                     270
PRO GLY PHE GLN THR PHE VAL SER CYS GLN GLU GLN ARG LEU THR TYR
            275                     280                     285
LEU PRO GLN PRO TRP GLY ASN CYS ARG ALA GLU SER GLU LEU ARG GLU
        290                     295                     300
PRO GLU LEU GLN GLY TYR SER ALA TYR SER VAL SER ALA CYS ARG LEU
305                     310                     315                     320
ARG CYS GLU LYS GLU ALA VAL XAA GLN ARG CYS HIS CYS ARG MET VAL
                        325                     330                     335
HIS MET PRO GLY ASN GLU THR ILE CYS PRO PRO ASN ILE TYR ILE GLU
                340                     345                     350
CYS ALA ASP XAA XAA XAA XAA XAA XAA XAA XAA XAA XAA XAA XAA XAA
            355                     360                     365
XAA XAA XAA XAA XAA XAA XAA XAA XAA XAA XAA XAA XAA XAA XAA XAA
        370                     375                     380
XAA XAA XAA XAA XAA XAA XAA XAA XAA XAA XAA XAA XAA XAA XAA XAA
385                     390                     395                     400
XAA XAA XAA XAA XAA XAA XAA XAA XAA XAA XAA XAA XAA XAA XAA XAA
                        405                     410                     415
XAA CYS PHE CYS PRO THR PRO CYS ASN LEU THR ARG TYR GLY LYS GLU
                420                     425                     430
ILE SER MET VAL ARG ILE PRO ASN ARG GLY SER ALA ARG TYR LEU THR
```

```
        435                 440                 445
ARG LYS TYR ASN ARG ASN GLU THR TYR ILE ARG GLU ASN PHE LEU VAL
    450                 455                 460
LEU ASP VAL PHE PHE GLU ALA LEU THR SER GLU ALA MET GLU GLN ARG
465                 470                 475                 480
ALA ALA TYR GLY LEU SER ALA LEU LEU GLY ASP LEU GLY GLY GLN MET
                485                 490                 495
GLY LEU PHE ILE GLY ALA SER ILE LEU THR LEU LEU GLU ILE LEU ASP
            500                 505                 510
TYR ILE TYR GLU VAL SER TRP ASP ARG LEU LYS ARG XAA MET GLU ALA
        515                 520                 525
SER GLN ASP PRO LEU ARG THR SER THR GLY GLY ILE SER THR LEU GLY
    530                 535                 540
LEU GLN GLU LEU LYS GLU GLN ILE PRO ALA XAA PRO GLY PRO XAA GLY
545                 550                 555                 560
ARG TRP GLY GLY GLN GLN SER ALA PRO ASN HIS HIS HIS PRO HIS GLY
                565                 570                 575
PRO PRO GLY SER LEU PHE GLU ASN PHE ALA CYS
            580                 585
```

(2) INFORMATION FOR SEQ ID NO:5:

(I) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 508 BASE PAIRS
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(II) MOLECULE TYPE: CDNA

(XI) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
AATTCGGCAC GAGTGCAACC TGACACGCTA TGGGNAAGAG ATCTCCATGG TCAGGATCCC    60
CAACAGGGGC TCAGCCCGGT ACCTGACGAG GAAGTACAAC CGCAACGAGA CCTACATACG   120
GGAGAACTTC CTGGTACTAG ATGTCTTCTT TGAGGCCCTG ACCTNTGAAG NCATGGAGCA   180
GCGAGCAGCC TATGGCCTGT CAGCCCTGCT GGGAGACCTC GGGGGACAGA TGGGNCTGTT   240
CATTGGGGCC AGCATCCTCA CGTTGCTGGA GATTCTTGAC TACATCTATG AGGTGTCCTG   300
GGGATCGACT TGAAGCGGGT ATTGGAGGCG TNCCNAGACC CCCCTTGNGG GACCTNCACT   360
TGGGGGGATT TCCAATTTTG GGGNTTCAAG AGGTTTNAGG AACANNNTTC CTNCCCNAGC   420
CGGGGCCCAT TTGAGGGTTG GGGGGTCAAA ANTTTTTTCC ATTAACACCA CCCCAGGTCC   480
CAGNAGTTTT TNAGGTTTTT TTTAGGNG                                      508
```

(2) INFORMATION FOR SEQ ID NO:6:

(I) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 103 AMINO ACIDS
(B) TYPE: AMINO ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(II) MOLECULE TYPE: PROTEIN

(XI) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
ILE ARG HIS GLU CYS ASN LEU THR ARG TYR GLY XAA GLU ILE SER MET
 1           5                   10                  15
VAL ARG ILE PRO ASN ARG GLY SER ALA ARG TYR LEU THR ARG LYS TYR
            20                  25                  30
ASN ARG ASN GLU THR TYR ILE ARG GLU ASN PHE LEU VAL LEU ASP VAL
        35                  40                  45
PHE PHE GLU ALA LEU THR XAA GLU XAA MET GLU GLN ARG ALA ALA TYR
    50                  55                  60
GLY LEU SER ALA LEU LEU GLY ASP LEU GLY GLY GLN MET GLY LEU PHE
65                  70                  75                  80
ILE GLY ALA SER ILE LEU THR LEU LEU GLU ILE LEU ASP TYR ILE TYR
                85                  90                  95
GLU VAL SER TRP GLY SER THR
            100
```

(2) INFORMATION FOR SEQ ID NO:7:

(I) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1632 BASE PAIRS
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(II) MOLECULE TYPE: CDNA

(XI) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
ATGCCGATCG AGATTGTGTG CAAAATCAAA TTTGCTGAGG AGGATGCGAA ACCCAAGGAG   60
AAGGAGGCAG GGGATGAGCA GAGCCTCCTC GGGGCTGTCG CCCCTGGAGC AGCCCCCCGA  120
GACCTGGCCA CCTTTGCCAG CACCAGCACC CTGCATGGAC TGGGCCGGGC CTGTGGCCCA  180
GGCCCCCACG GACTGCGCAG AACCCTGTGG GCACTGGCCC TACTCACCTC GCTGGCTGCC  240
TTCCTGTACC AGGCGGCTGG CCTGGCCCGG GGCTACCTGA CCCGGCCTCA CCTGGTGGCA  300
ATGGACCCCG CTGCCCCAGC CCCAGTGGCG GGCTTCCCGG CTGTCACCCT CTGCAATATC  360
AACCGCTTCC GGCATTCGGC ACTCAGCGAT GCCGACATCT TCCACCTGGC CAATCTGACA  420
GGGCTGCCCC CCAAAGACCG GGATGGGCAC CGTGCGGCTG GCCTGCGCTA CCCAGAGCCT  480
GACATGGTAG ACATCCTCAA CCGCACTGGC CACCAGCTCG CCGACATGCT TAAGAGCTGC  540
```

```
AACTTCAGTG GGCATCACTG CTCCGCCAGC AACTTCTCTG TGGTCTATAC TCGCTATGGG      600
AAGTGTTACA CCTTCAACGC GGACCCGCGG AGCTCGCTGC CCAGCCGGGC AGGGGGCATG      660
GGCAGTGGCC TGGAGATCAT GCTGGACATC CAGCAGGAGG AGTACCTGCC CATCTGGAGG      720
GAGACAAATG AGACGTCGTT TGAGGCAGGT ATTCGGGTGC AGATCCACAG CCAGGAGGAG      780
CCGCCCTACA TCCACCAGCT GGGGTTCGGG GTGTCCCCAG GCTTCCAGAC CTTTGTGTCC      840
TGCCAGGAAC AGCGGCTGAC CTACCTGCCC CAGCCCTGGG GCAACTGCCG CGCAGAGAGT      900
GAGCTCAGGG AGCCTGAGCT TCAGGGCTAC TCGGCCTACA GTGTGTCTGC CTGCCGGCTG      960
CGCTGTGAAA AGGAGGCCGT GCTTCAGCGC TGCCACTGCC GGATGGTGCA CATGCCAGGC     1020
AATGAGACCA TCTGCCCACC AAATATCTAC ATCGAGTGTG CAGACCACAC ACTGGACTCC     1080
CTGGGTGGGG GCCCTGAGGG CCCGTGCTTC TGCCCCACCC CCTGCAACCT GACACGCTAT     1140
GGGAAAGAGA TCTCCATGGT CAGGATCCCC AACAGGGGCT CAGCCCGGTA CCTGGCGAGG     1200
AAGTACAACC GCAACGAGAC CTACATACGG GAGAACTTCC TGGTCCTAGA TGTCTTCTTT     1260
GAGGCCCTGA CCTCTGAAGC CATGGAGCAG CGAGCAGCCT ATGGCCTGTC AGCCCTGCTG     1320
GGAGACCTCG GGGGACAGAT GGGCCTGTTC ATTGGGGCCA GCATCCTCAC GTTGCTGGAG     1380
ATCCTCGACT ACATCTATGA GGTGTCCTGG GATCGACTGA AGCGGGTATG GAGGCGTCCC     1440
AAGACCCCCC TGCGGACCTC CACTGGGGGC ATCTCCACTT TGGGGCTTCA GGAGCTGAAG     1500
GAACAGAGTC CCTGTCCGAG CCTGGGCCGA GCGGAGGGTG GGGGGGTCAG CAGTCTGCTC     1560
CCCAATCACC ACCACCCCCA CGGTCCCCCA GGAGGTCTCT TTGAAGATTT TGCTTGCTAG     1620
GACGGTGCTG TG                                                         1632
```

(2) INFORMATION FOR SEQ ID NO:8:

(I) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 539 AMINO ACIDS

(B) TYPE: AMINO ACID

(C) STRANDEDNESS: SINGLE

(D) TOPOLOGY: LINEAR

(II) MOLECULE TYPE: PROTEIN

(XI) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
MET PRO ILE GLU ILE VAL CYS LYS ILE LYS PHE ALA GLU GLU ASP ALA
 1               5                  10                  15
LYS PRO LYS GLU LYS GLU ALA GLY ASP GLU GLN SER LEU LEU GLY ALA
            20                  25                  30
VAL ALA PRO GLY ALA ALA PRO ARG ASP LEU ALA THR PHE ALA SER THR
        35                  40                  45
SER THR LEU HIS GLY LEU GLY ARG ALA CYS GLY PRO GLY PRO HIS GLY
    50                  55                  60
LEU ARG ARG THR LEU TRP ALA LEU ALA LEU LEU THR SER LEU ALA ALA
65                  70                  75                  80
PHE LEU TYR GLN ALA ALA GLY LEU ALA ARG GLY TYR LEU THR ARG PRO
                85                  90                  95
HIS LEU VAL ALA MET ASP PRO ALA ALA PRO ALA PRO VAL ALA GLY PHE
            100                 105                 110
```

PRO ALA VAL THR LEU CYS ASN ILE ASN ARG PHE ARG HIS SER ALA LEU
115 120 125

SER ASP ALA ASP ILE PHE HIS LEU ALA ASN LEU THR GLY LEU PRO PRO
130 135 140

LYS ASP ARG ASP GLY HIS ARG ALA ALA GLY LEU ARG TYR PRO GLU PRO
145 150 155 160

ASP MET VAL ASP ILE LEU ASN ARG THR GLY HIS GLN LEU ALA ASP MET
165 170 175

LEU LYS SER CYS ASN PHE SER GLY HIS HIS CYS SER ALA SER ASN PHE
180 185 190

SER VAL VAL TYR THR ARG TYR GLY LYS CYS TYR THR PHE ASN ALA ASP
195 200 205

PRO ARG SER SER LEU PRO SER ARG ALA GLY GLY MET GLY SER GLY LEU
210 215 220

GLU ILE MET LEU ASP ILE GLN GLN GLU GLU TYR LEU PRO ILE TRP ARG
225 230 235 240

GLU THR ASN GLU THR SER PHE GLU ALA GLY ILE ARG VAL GLN ILE HIS
245 250 255

SER GLN GLU GLU PRO PRO TYR ILE HIS GLN LEU GLY PHE GLY VAL SER
260 265 270

PRO GLY PHE GLN THR PHE VAL SER CYS GLN GLU GLN ARG LEU THR TYR
275 280 285

LEU PRO GLN PRO TRP GLY ASN CYS ARG ALA GLU SER GLU LEU ARG GLU
290 295 300

PRO GLU LEU GLN GLY TYR SER ALA TYR SER VAL SER ALA CYS ARG LEU
305 310 315 320

ARG CYS GLU LYS GLU ALA VAL LEU GLN ARG CYS HIS CYS ARG MET VAL
325 330 335

HIS MET PRO GLY ASN GLU THR ILE CYS PRO PRO ASN ILE TYR ILE GLU
340 345 350

CYS ALA ASP HIS THR LEU ASP SER LEU GLY GLY GLY PRO GLU GLY PRO
355 360 365

CYS PHE CYS PRO THR PRO CYS ASN LEU THR ARG TYR GLY LYS GLU ILE
370 375 380

SER MET VAL ARG ILE PRO ASN ARG GLY SER ALA ARG TYR LEU ALA ARG
385 390 395 400

LYS TYR ASN ARG ASN GLU THR TYR ILE ARG GLU ASN PHE LEU VAL LEU
405 410 415

ASP VAL PHE PHE GLU ALA LEU THR SER GLU ALA MET GLU GLN ARG ALA
420 425 430

ALA TYR GLY LEU SER ALA LEU LEU GLY ASP LEU GLY GLY GLN MET GLY
435 440 445

LEU PHE ILE GLY ALA SER ILE LEU THR LEU LEU GLU ILE LEU ASP TYR
450 455 460

ILE	TYR	GLU	VAL	SER	TRP	ASP	ARG	LEU	LYS	ARG	VAL	TRP	ARG	ARG	PRO
465					470					475					480
LYS	THR	PRO	LEU	ARG	THR	SER	THR	GLY	GLY	ILE	SER	THR	LEU	GLY	LEU
				485					490					495	
GLN	GLU	LEU	LYS	GLU	GLN	SER	PRO	CYS	PRO	SER	LEU	GLY	ARG	ALA	GLU
			500					505					510		
GLY	GLY	GLY	VAL	SER	SER	LEU	LEU	PRO	ASN	HIS	HIS	HIS	PRO	HIS	GLY
		515					520					525			
PRO	PRO	GLY	GLY	LEU	PHE	GLU	ASP	PHE	ALA	CYS					
	530					535									

Claims

1. An isolated polypeptide selected from:

   (a) a polypeptide comprising an amino acid sequence which has at least 70% identity to the amino acid sequence of SEQ ID NO:2 and
   (b) a polypeptide comprising an amino acid sequence which has at least 70% identity to the amino acid sequence of SEQ ID NO:8.

2. An isolated polypeptide as claimed in claim 1 in which the amino acid sequence has at least 95% identity.

3. The polypeptide as claimed in claim 1 comprising the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:8.

4. The isolated polypeptide of SEQ ID NO:2 or SEQ ID NO:8.

5. An isolated polynucleotide selected from:

   (a) a polynucleotide comprising a nucleotide sequence encoding a polypeptide that has at least 70% identity to the amino acid sequence of SEQ ID NO:2 over the entire length of SEQ ID NO:2, and
   (b) a polynucleotide comprising a nucleotide sequence encoding a polypeptide that has at least 70% identity to the amino acid sequence of SEQ ID NO:8 over the entire length of SEQ ID NO:8,

   or a nucleotide sequence complementary to said isolated polynucleotide.

6. An isolated polynucleotide selected from:

   (a) a polynucleotide comprising a nucleotide sequence that has at least 75% identity to a nucleotide sequence encoding a polypeptide of SEQ ID NO:2 over the entire coding region, and
   (b) a a polynucleotide comprising a nucleotide sequence that has at least 75% identity to a nucleotide sequence encoding a polypeptide of SEQ ID NO:8, over the entire coding region,

   or a nucleotide sequence complementary to said isolated polynucleotide.

7. An isolated polynucleotide selected from:

   (a) a polynucleotide which comprises a nucleotide sequence which has at least 75% identity to that of SEQ ID NO: 1 over the entire length of SEQ ID NO:1, and
   (b) a polynucleotide which comprises a nucleotide sequence which has at least 75% identity to that of SEQ ID NO:7 over the entire length of SEQ ID NO:7,

or a nucleotide sequence complementary to said isolated polynucleotide.

8. The isolated polynucleotide as claimed in any one of claims 5 to 7 in which the identity is at least 95%.

9. An isolated polynucleotide selected from:

(a) a polynucleotide comprising a nucleotide sequence encoding the polypeptide of SEQ ID NO:2,
(b) the polynucleotide of SEQ ID NO: 1:
(c) a polynucleotide obtainable by screening an appropriate library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO:1 or a fragment thereof, or a nucleotide sequence complementary to said isolated polynucleotide.

10. An isolated polynucleotide selected from:

(a) a polynucleotide comprising a nucleotide sequence encoding the polypeptide of SEQ ID NO:8,
(b) the polynucleotide of SEQ ID NO:7:
(c) a polynucleotide obtainable by screening an appropriate library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO:7 or a fragment thereof.

or a nucleotide sequence complementary to said isolated polynucleotide.

11. An expression system comprising a polynucleotide capable of producing a polypeptide of claim 1 when said expression system is present in a compatible host cell.

12. A host cell comprising the expression system of claim 11 or a membrane thereof expressing the polypeptide of claim 1.

13. A process for producing a polypeptide of claim 1 comprising culturing a host cell of claim 12 under conditions sufficient for the production of said polypeptide and recovering the polypeptide from the culture medium.

14. An antibody immunospecific for the polypeptide of claim 1.

15. A method for screening to identify compounds which stimulate or which inhibit the function of the polypeptide of claim 1 which comprises a method selected from the group consisting of:

(a) measuring the binding of a candidate compound to the polypeptide (or to the cells or membranes bearing the polypeptide) or a fusion protein thereof by means of a label directly or indirectly associated with the candidate compound:
(b) measuring the binding of a candidate compound to the polypeptide (or to the cells or membranes bearing the polypeptide) or a fusion protein thereof in the presense of a labeled competitior;
(c) testing whether the candidate compound results in a signal generated by activation or inhibition of the polypeptide, using detection systems appropriate to the cells or cell membranes bearing the polypeptide:
(d) mixing a candidate compound with a solution containing a polypeptide of claim 1, to form a mixture, measuring activity of the polypeptide in the mixture, and comparing the activity of the mixture to a standard: or
(e) detecting the effect of a candidate compound on the production of mRNA encoding said polypeptide and said polypeptide in cells, using for instance, an ELISA assay.

16. An agonist or antagonist to the polypeptide of claims 1 to 4.

17. A compound which is:

(a) an agonist or antagonist to the polypeptide of claims 1 to 4;
(b) isolated polynucleotide of claims 5 to 10; or
(c) a nucleic acid molecule that modulates the expression of the nucleotide sequence encoding the polypeptide of claim 1;

for use in therapy.

18. A process for diagnosing a disease or a susceptibility to a disease in a subject related to expression or activity of the polypeptide of claim 1 in a subject comprising:

(a) determining the presence or absence of a mutation in the nucleotide sequence encoding said polypeptide in the genome of said subject; and/or
(b) analyzing for the presence or amount of said polypeptide expression in a sample derived from said subject.

19. An isolated polynucleotide selected form the group consisting of:

(a) an isolated polynucleotide comprising a nucleotide sequence which has at least 75% identity to SEQ ID NO:3 over the entire length of SEQ ID NO:3;
(b) an isolated polynucleotide comprising the polynucleotide of SEQ ID NO:3;
(c) the polynucleotide of SEQ ID NO:3; or
(d) an isolated polynucleotide comprising a nucleotide sequence encoding a polypeptide which has at least 70% identity to the amino acid sequence of SEQ ID NO:4, over the entire length of SEQ ID NO:4.

20. A polypeptide selected from the group consisting of:

(a) a polypeptide which comprises an amino acid sequence which has at least 70% identity to that of SEQ ID NO:4 over the entire length of SEQ ID NO:4;
(b) a polypeptide in which the amino acid sequence has at least 70% identity to the amino acid sequence of SEQ ID NO:4 over the entire length of SEQ ID NO:4;
(c) a polypeptide which comprises the amino acid of SEQ ID NO:4;
(d) a polypeptide which is the polypeptide of SEQ ID NO:4; or
(e) a polypeptide which is encoded by a polynucleotide comprising the sequence contained in SEQ ID NO:3.

21. An isolated polynucleotide selected form the group consisting of:

(a) an isolated polynucleotide comprising a nucleotide sequence which has at least 80% identity to SEQ ID NO:5 over the entire length of SEQ ID NO:5;
(b) an isolated polynucleotide comprising the polynucleotide of SEQ ID NO:5;
(c) the polynucleotide of SEQ ID NO:5; or
(d) an isolated polynucleotide comprising a nucleotide sequence encoding a polypeptide which has at least 70% identity to the amino acid sequence of SEQ ID NO:6, over the entire length of SEQ ID NO:6.

22. A polypeptide selected from the group consisting of:

(a) a polypeptide which comprises an amino acid sequence which has at least 70% identity to that of SEQ ID NO:6 over the entire length of SEQ ID NO:6;
(b) a polypeptide in which the amino acid sequence has at least 70% identity to the amino acid sequence of SEQ ID NO:6 over the entire length of SEQ ID NO:6;
(c) a polypeptide which comprises the amino acid of SEQ ID NO:6;
(d) a polypeptide which is the polypeptide of SEQ ID NO:6; or
(e) a polypeptide which is encoded by a polynucleotide comprising the sequence contained in SEQ ID NO:5.